PB88-108675

Environmental Monitoring and Enforcement Theory and Practice

Association of Environmental and Resource
Economists, Washington, DC

Prepared for

Environmental Protection Agency, Washington, DC

Aug 87

U.S. Department of Commerce
National Technical Information Service
NTIS

PB88-108675

EPA-230-10-87-030

ASSOCIATION OF ENVIRONMENTAL AND RESOURCE ECONOMISTS (AERE)

WORKSHOP

ENVIRONMENTAL MONITORING AND ENFORCEMENT:
THEORY AND PRACTICE

University of Delaware
Clayton Hall
Newark, Delaware 19716

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

August 13-14, 1987

The information in this document has been funded wholly or in part by the United States Envirnomental Protection Agency under Cooperative Agreement No. CR-812056 to the Association of Environmental and Resource Economists. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

50272-101

REPORT DOCUMENTATION PAGE	**1. REPORT NO.** EPA 230-10-87-030	**2.**	**3. Recipient's Accession No.** PB88-108675/AS

4. Title and Subtitle

Environmental Monitoring and Enforcement: Theory and Practice

5. Report Date

August 1987

6.

7. Author(s)

Multiple Authors

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

Association of Environmental and Resource Economists
1616 P Street, N.W.
Washington, D.C. 20036

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.

(C)

(G) CR812056

12. Sponsoring Organization Name and Address

Office of Policy Analysis, PM-221
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, D.C. 20460

The report

13. Type of Report & Period Covered

Final

14.

15. Supplementary Notes

Reports from two earlier workshops -- Recreation Demand Modeling (NTIS #PB86-109642) and Marine Pollution and Environmental Damage Assessment (also being sent to NTIS) are available.

16. Abstract (Limit: 200 words)

Contains 9 papers prepared for the August 13-14, 1987 workshop sponsored by the Association of Environmental and Resource Economists (AERE) on the subject of Environmental Monitoring and Enforcement. Brief titles are given below:

C. Wasserman, "Environmental Compliance and Enforcement;" (L. G. Anderson, "Enforcement Issues in Selecting Policy Instruments;") C. S. Russell, "Game Theory Lessons for Structuring Montoring and Enforcement Systems;" G. V. Johnson and T. S. Ulen, "The Behavior of the Firm when Facing Uncertain Enforcement;" A. S. Malik, "Markets for Pollution Control when Firms are Noncompliant;" J. G. Sutinen and J. R. Gauvin, "Enforcement and Compliance in the Massachusetts Lobster Fishery;" W. Gray and C. A. Jones, "Impact of OSHA Regulations in Manufacturing;" R. Rosenman, "The Regulatory Loss of Soil Management Policy;" W. A. Magat and W. K. Viscusi, "Enforcement: Industrial Effluent Standards."

17. Document Analysis a. Descriptors

Environmental monitoring
Environmental enforcement

b. Identifiers/Open-Ended Terms

compliance with environmental controls

c. COSATI Field/Group

18. Availability Statement: No restriction on distribution. Available from NTIS, Springfield, VA 22161	**19. Security Class (This Report)** Unclassified	**21. No. of Pages**
	20. Security Class (This Page) Unclassified	**22. Price**

(See ANSI–Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4–77)
(Formerly NTIS–35)
Department of Commerce

ASSOCIATION OF ENVIRONMENTAL AND RESOURCE ECONOMISTS (AERE)

WORKSHOP

ENVIRONMENTAL MONITORING AND ENFORCEMENT: THEORY AND PRACTICE

SESSION I

Agency and Judicial Perspectives on Environmental Enforcement Policies

Cheryl Wasserman, U.S. Environmental Protection Agency
"Environmental Compliance and Enforcement: Theory, Practice, and the Challenge to Environmental Economists"

Nancy Firestone, U.S. Department of Justice
"An Overview of Natural Resource Enforcement Issues at the Department of Justice" (No paper accompanying presentation)

Lee G. Anderson, University of Delaware
"Enforcement Issues in Selecting Policy Instruments"

SESSION II

Modeling Optimal Enforcement Strategies

Clifford S. Russell, Vanderbilt University
"Game Theory Lessons for Structuring Monitoring and Enforcement Systems"
Clifford S. Russell, Vanderbilt University

Gary V. Johnson and Thomas S. Ulen, University of Illinois
"The Behavior of the Firm When Facing Uncertain Enforcement of Ex Ante Environmental Regulation and Tort Liability Standards"

Arun S. Malik, University of Maryland
"Markets for Pollution Control When Firms are Noncompliant"

SESSION III

Empirical Models of Compliance and Performance of Enforcement Agencies

Jon G. Sutinen and John R. Gauvin, University of Rhode Island
"Enforcement and Compliance in the Massachusetts Lobster Fishery"

Wayne Gray, Clark University and Carol A. Jones, University of Michigan
"Longitudinal Analysis of the Impact of OSHA Health and Safety Regulations in Manufacturing"

Robert Rosenman, Washington State University
"Hitting the 'T': The Regulatory Loss of Soil Management Policy"

Wesley A. Magat, Duke University and W. Kip Viscusi, Northwestern University
"The Effectiveness of EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards"

SESSION I

Agency and Judicial Perspectives on Environmental Enforcement Policies

Cheryl Wasserman, U.S. Environmental Protection Agency
"Environmental Compliance and Enforcement: Theory, Practice, and the Challenge to Environmental Economists"

Nancy Firestone, U.S. Department of Justice
"An Overview of Natural Resource Enforcement Issues at the Department of Justice" (No paper accompanying presentation)

Lee G. Anderson, University of Delaware
"Enforcement Issues in Selecting Policy Instruments"

"ENVIRONMENTAL COMPLIANCE AND ENFORCEMENT:
THEORY, PRACTICE, AND THE CHALLENGE TO ENVIRONMENTAL ECONOMISTS"

Delivered to

Association of Environmental and Resource Economists (AERE)

Workshop on

Environmental Monitoring and Enforcement:
Theory and Practice

at

University of Delaware
Newark, Delaware 19716

Prepared by

Cheryl Wasserman
Chief, Compliance Policy and Planning
Office of Enforcement and Compliance Monitoring
U.S. Environmental Protection Agency

July 14, 1987

ENVIRONMENTAL COMPLIANCE AND ENFORCEMENT: THEORY, PRACTICE, AND THE CHALLENGE TO ENVIRONMENTAL ECONOMISTS

I. INTRODUCTION

Compliance is essential to the success of any environmental program. However, there is little empirical data on how best to gain broad-scale compliance with environmental laws. There is comparatively little in the traditional economics literature on the subject of compliance, and still less on environmental compliance and enforcement: what works, does not work, in what circumstances, and why. This paper reviews the theories that federal, state and local regulators and law enforcement personnel draw upon in implementing enforcement programs, how the realities of implementation differ from theory, and the need for further research by environmental economists and those in related social science disciplines.

II. WHY COMPLIANCE AND ENFORCEMENT MATTER

In the early 1970s much of the attention of government officials at the federal, state, and local levels was devoted to standard setting, and to permit issuance. While enforcement was undertaken as a basic element of each environmental program, it first began to assume its place as the subject of significiant policy interest and priority in the latter part of the 1970s and 1980s. This in large part reflects a natural evolution in the programs, shifting to field implementation once requirements were put in place and in part, reflects the introduction of statutory penalty authorities in the latter part of the 1970s. There are several reasons why concerns about the extent of compliance and the effectiveness of enforcement efforts are important:

1. Effectiveness: Compliance is critical to realizing the benefits envisioned by environmental policy, statutes, regulations, standards, and permits. The vast regulatory apparatus we have put in place to protect public health and the environment amounts to empty words and deeds without compliance. It is the regulatory bottom line. All of the cost-benefit or cost-effectiveness studies that may go into assessing the best regulatory route become meaningless if the costs and benefits do not play out as predicted, usually with assumptions of full compliance.

2. Efficiency: Environmental economics, among other things, seeks to identify the most efficient solution to achieving desired public health and welfare benefits. If regulations and permit conditions are designed to be economically efficient, inconsistent enforcement will lead to economically inefficient results.1/

3. Equity: A consistent enforcement response provides an element of fairness to the regulatory process that would be missing if those who failed to comply benefit relative to those who do, or if regulatees are treated very differently depending upon their location and circumstance.

4. Credibility: The rule of law and the credibility of our governmental institutions require that laws that are put in force are taken seriously. The expectation that violations will generate a predictable and proportionate enforcement response is essential to the credibility of our regulations. Support for escalating enforce-

ment response to known violations, and following through from the less costly to the more resource intensive enforcement responses until compliance is achieved is the ultimate test of the public will to see a program through to its full implementation.

III. BASIC THEORIES ON COMPLIANCE AND ENFORCEMENT

There are several concepts and/or operative theories underlying environmental enforcement; some are explicit in guidance and policy, while others can be surmised from practice through observation. All are interrelated. This discussion first distinguishes between compliance and enforcement. Second, it presents the theory of deterrence fundamental to enforcement strategies and the focus of the contributions of economic theory to date. Third, it discusses other theories or assumptions about what motivates compliance behavior, and the implications for more efficient and effective compliance and enforcement programs.

Compliance is defined as the ultimate goal of any enforcement program. Compliance is essentially a state of being. When a regulated source is in compliance, it is achieving required environmental standards or regulations. Enforcement is defined as a set of legal tools, both informal and formal, designed to achieve compliance. The term enforcement response usually applies to those agency actions specifically intended to convey legal sanction and/or penalty. Traditional enforcement programs only encompass compliance monitoring and informal and formal enforcement response. Recently, the broader range of both coercive and noncoercive techniques used to effectively change or deter source behavior are being tied together into comprehensive compliance strategies for each of the major environmental programs and their components.

Deterrence is perhaps the most important underlying theory of enforcement; however, it is used somewhat differently in the literature than in practice. The concept of deterrence is that a strong enforcement program deters the regulated community from violating in the first instance. Specific deterrence is deterring an individual violator from violating again. General deterrence is deterring the broader regulated community from violating. The prevailing theory on creating deterrence states that three elements are necessary: 1) a credible likelihood of detection of the violation; 2) swift and sure enforcement response; and 3) appropriately severe sanction.2/ Some would add the condition that each of these factors be perceived as real, whether supported or not.

Deterrence is viewed in practice as creating a multiplier effect for each enforcement action, the magnitude of which depends on the strength of each of these factors. The multiplier effect is important. No enforcement program can provide sufficient presence all of the time for all violations. It is therefore generally held that an enforcement program must rely upon and to some extent develop a complying majority and devote its resources to addressing the remainder that do not comply. In contrast, pure deterrence theory as described in the literature would have us believe that there is no "voluntary compliance", that it is only the direct threat of some legal action and sanction that motivates a source to comply.

Economic theory as it has been applied to enforcement is compatible with and indeed could be said to quantify the principles of deterrence theory. Both describe the would be-violator as an amoral, rational calculator. Economic theory argues that each source calculates whether it is to its economic self interest to comply or to violate requirements. This economic calculus compares the cost of coming into compliance with the likelihood of getting caught times the penalty of violating. Economic self-interest could also take into account future liability, the economic value of a negative impact on the firm°s reputation if caught, etc., but at bottom it comes down to dollars and cents.

Neither pure deterrence theory nor its economic compliment addresses the broader array of motivations that may explain compliance behavior, such as societal norms, moral values, sense of professional conduct, etc. Because compliance behavior is just that, behavior, it cannot readily be predicted and is more likely the result of complex motivations, only some of which are rational, reasonable, and economically motivated. That is not to say that economic theory and analysis is without merit or potential value here. It is to say that environmental economics as applied to enforcement is not a simple matter.

The behavioral school of compliance theory argues that, at least for corporate compliance, individuals within a firm are motivated less by conscious decisions based on profit/loss less than motives of personal advancement, by fear of corporate sanction, or by social influence through an individual relationship with the regulator/inspector, peers, and/or social and moral norms. 4/ The literature takes two approaches to this more behavior-driven compliance. One group presents the business firm or regulatee as a political citizen, postulating that enforcement is not costeffective because of its high transaction costs, and that it is really not necessary in most instances given the inherent willingness of sources to comply with the law. This school argues for a more cooperative approach to gaining compliance, and to the promulgation of rules that are perceived as reasonable, since those motivated by citizenship will not obey rules perceived as arbitrary or irrational.

Another group in the literature focuses less on cooperation than on sanctions, arguing that because fines and penalties are unlikely to be sufficiently high to overcome the low likelihood of detection and are not coming directly out of the individual°s pocket, that criminal sanctions, with the threat of personal incarceration, or other sanctions adversely affecting individuals within an organization, are far more powerful tools to gain compliance. 5/

The literature presents a third potential image of the violator, that of the incompetent, with violations stemming from a failed organization or procedure. Under deterrence theory, the threat of sanction is thought to be the best motivation to establish proper systems and internal controls to motivate against the safety of blissful ignorance. Those who advocate for cooperation, argue that a strong enforcement response is inappropriate here as it is the lack of capacity, and not the will to comply that must be overcome. As to what truly motivates preventive behavior, there will soon be an emerging body of literature on corporate culture and how it might affect compliance behavior. Some case studies are under development by Tufts University. EPA is funding this project as well as a series of expert groups from academia, industry, and government regulators on what motivates compliance. The bulk of this work is being done through the University of Buffalo School of Law. 6/

Some of the recent work in the field of enforcement is derived from game theory and operations research. This work attempts to establish a research framework for predicting behavior, given the complex relationship between the regulator and regulatee, borrowing from economic literature but adding situational variables that respond more accurately to a variety of possible motivations to comply. 7/

One factor to keep in mind is that, in enforcement, perception is as important or even more important than reality. The perception of a strong enforcement effort, of the willingness to pursue sanctions and to escalate enforcement action can be created by the manner in which government enforcement actions are taken, as well as by the actual number of actions themselves. While one cannot create perceived action from nonaction, in the annals of military history, there have been many successful battles won where few in number have created an image of a formidable fighting force, thereby demoralizing and successfully competing against superior forces. The point is that there is an element to a successful enforcement effort that cannot be correlated to quantification of action and result; there is a multiplier effect that must capture these other factors or theory is likely to paint an unrealistically gloomy portrait of what it would take to do the job.

Finally, enforcement strategies must reflect a mix of these theories. At any stage in the process*, the mix of compliance promotion and enforcement response could vary, based upon the nature of the requirements, the level of public support and technical know-how. Moreover, for students of human nature, it would seem, that there is always a need for some enforcement. Chester Bowles, in 1971 of the wartime Office of Price Administration suggested that there will always be 5% of individuals who will violate no matter what, 20% who will comply no matter what, and 75% who will comply only if the violators are punished and/or the requirements are perceived as nonarbitrary. This means that there will always be a question of timing and mix of enforcement response versus technical assistance, whether you "stroke°em" or "poke°em" first or second or at the same time, in some changing proportion.

---

*It may be useful to postulate five stages for compliance and enforcement programs: Stage I, the compliance promotion phase, is primarily cooperative. Regulated entities need clear communications on what is required and how they might go about complying with requirements. This stage occurs during the time lag before sources must comply with new requirements. Stage II might be called the enforcement initiative phase, at which time it might be most important for enforcement to be visible and forceful in initiating the program's requirements, and in creating the public's perception of their importance. At this stage, visible enforcement need not be extensive and might best be accompanied by continued technical assistance. Stage III, the initial compliance stage, during which the installation of control or monitoring equipment and the adoption of certain procedures, reports or practices is required, focuses on whether or not these items are in place. This stage may require more extensive enforcement to systematically ensure compliance. Stage IV, Continuing Compliance, emphasizes operations and maintenance or management practices and it may suffice to use selective enforcement combined with information dissemination. Finally, Stage V, could represent steady state compliance, or a return to stage III for retrofitting aging equipment.

IV. COMPLIANCE MONITORING AND ENFORCEMENT PROGRAMS: THEORY, REALITY, AND OPEN QUESTIONS

The strategies currently in use at the U.S. Environmental Protection Agency and its counterparts at the State and local levels generally address the following elements:

- establishing program priorities and identifying the regulated community;
- promoting compliance within the regulated community;
- monitoring compliance within the regulated community;
- responding to violations including timeliness and appropriateness of the enforcement response;
- imposing civil penalties and other sanctions for non-compliance;
- clarifying roles and responsibilities of federal and state or local agencies; and
- evaluating results.

These elements of a compliance strategy are reviewed below in terms of the current operative theory used by government officials, what we know of practice and how it differs from theory, and finally, what questions should be addressed in further research to lead to the design of more cost-effective compliance and enforcement programs.

1. Establishing program priorities and identifying the regulated community

Environmental requirements now cover virtually every activity involving production, transportation and consumption in our society. Keeping up with the most recent developments in environmental policy is more than a full-time task. The number of Federal Register pages devoted to environmental requirements have grown geometrically, and plant-specific permit requirements are now reflecting the complete environmental impact cycle from storage of raw materials of production to disposal of process waste products. As difficult as it is for the regulated community to keep up with these requirements, it is even more difficult for environmental officials to ensure compliance is achieved and to take the necessary enforcement actions. Therefore, priorities must be set to focus enforcement activities.

In theory, enforcement priorities should simply take into account the health and environmental risk posed by failure to comply, much as we do when devising standards and regulations. However, enforcement poses more complex challenges for priority setting. Some requirements, if violated, may pose a very high environmental risk but low probability of violation, such as a requirement where there is liability for damages through other institutions. This might arguably make it a low enforcement priority. Also a violation may be one which has a very low risk of occuring given the reliability of the control system, or may be one for which only a small percentage of the regulated community is out of compliance.

Some requirements for which the violation does not pose any potential threat or harm to public health or to the environment, such as some monitoring, reporting, and record keeping requirements may be essential to the integrity of the regulatory scheme. In the case of self-reported compliance monitoring data, it might be essential to detecting when and where more serious violations and harm may occur. Similarly, a violation of a legal order or consent decree may not in itself be significant but must be enforced to demonstrate the significance of a legally imposed order. Therefore, enforcement priorities must reflect an effective mix of concerns for the risk of environmental harm, and the need for a credible enforcement presence.

Enforcement priorities are currently defined through several vehicles. At a very broad level, annual Agency Guidance specifies program priorities. EPA has placed a premium on fully integrating enforcement into every aspect of program implementation. For example, if a major thrust of the Toxics Program is pre-manufacture notification of new chemicals, there must be a parallel enforcement component to that priority which ensures that industry does indeed perform the necessary testing and report results of all of the appropriate test data. More specfically, enforcement priorities are further established in program measures of success which accompany the annual guidance, defining what constitutes Significant Non-Compliance (SNC) in each program. 8/ A measure of program success is then how effectively the Significant Non-Compliers are returned to compliance. In addition, there are other program areas for which an enforcement presence is needed but not necessarily through coverage of all violations. In such aspects of the program, enforcement initiatives may be designed to send a clear message to the regulated community. Enforcement initiatives are targeted and concentrated enforcement actions are timed to have a maximum deterrent impact through press coverage, and packaged to gain economies of scale in preparing cases for litigation.

In practice, the environmental programs have not been entirely successful in effectively establishing priorities based upon risk or on the need for deterrence. In the hazardous waste program for example, an effort was made to establish enforcement priorities at land disposal facilities for groundwater related violations. However, in reality, all groundwater violations do not pose the same level of risk and some very minor violations end up being included on the Significant Non-Complier (SNC) list. The Air program has placed a high priority on violations of pollutant standards in areas exceeding national ambient air quality standards for that pollutant. However, the air program also has included as SNC, violations of any national new source performance standard. Therefore, the failure to conduct a performance stack test upon starting up a new source in an attainment area, has had the same priority as a violation of volatile organic compound (VOC) requirements in an ozone nonattainment area even if federal or state officials have reason to believe the source is in compliance. While refinements are being made, a priority setting scheme is never fully satisfactory at the national level. Federal officials are currently working with EPA Regions and the States to more successfully adjust national priorities to local circumstances.

In a more detailed setting, priorities are also established in those programs which must address continuing compliance and operations and maintainance violations, e.g., air and water. Because no pollution control system can operate in one hundred percent compliance, one hundred percent of the time, these programs establish priorities by setting limits on actionable violation levels. This does not mean that any violations are countenanced but rather, that given the large numbers of violating sources, limits must be set. Because these limits are for management purposes only, they do not apply to citizen suits, and citizen suits have been effectively brought for violations at lesser levels which were not being addressed by regulatory agencies.

Further, systems are being put in place now within the U.S. EPA to assess whether enforcement actions taken are indeed those that have strategic value and the greatest return for the enforcement investment. All indications are that the majority of enforcement actions are focused on the most significant problems. Where they are not, oversight systems will attempt to correct future case selection.

In reviewing enforcement priorities it is important to recognize that an enforcement response happens late in the enforcement process. Faced with violations, most government agencies would prefer to respond in some manner. Therefore, strategies for detecting violations and for setting correct priorities at the beginning of the enforcement process can best assure that agency resources are focused on the most important problems. This issue is addressed below in Section 3. Monitoring Compliance.

2. Promoting Compliance within the Regulated Community

Compliance promotion is that set of activities which promote rather than coerce compliance, such as offers of assistance, information exchange, and the like. The theory that underlies most compliance monitoring and enforcement programs is that the best way to promote compliance is to enforce the law; a corollary is that ignorance is no excuse under the law. Nevertheless, in operation, compliance programs at the federal and state level offer a range of programs to disseminate information and provide technical assistance to the regulated community. While it is broadly held that the threat of enforcement is the best motivation for regulatees to avail themselves of these sources of information and assistance, state and federal officials do undertake these compliance promotion activities with a view toward building a complying majority, and tapping the broadest range of motivations to comply.

In practice, compliance promotion activities have been the least well funded and the most expendable activity in compliance and enforcement programs, although some would argue that each time an inspector visits a facility it is and should be used as an opportunity to inform, offer assistance, and convey credibility to the regulatory scheme. In recent years, however, with regulatory activities reaching ever smaller and more numerous sources, providing information to the regulated community (i.e. on the requirements for compliance, on why the requirements are important, on what is required to comply and the consequences of noncompliance) is viewed as more essential by federal and state regulators. A notable exception reflecting a trend in the opposite direction is the fact that until quite recently compliance promotion activities were the primary response to municipalities for safe drinking water and municipal treatment plant operating violations. The low levels of enforcement against these sources has not been effective and EPA has shifted its emphasis to introduce more vigorous enforcement action. In the case of safe drinking water, Congress has also had a hand in this transition by requiring, in the recent Amendments, enforcement for all violations, a goal which both EPA and States will have difficulty fulfilling.

One research issue that emerges is the efficiency of dollars spent promoting compliance versus enforcing requirements, and the proper balance between the two. Other issues concern the form and nature of regulations themselves as to how likely it is to elicit compliance behavior with little or no help from enforcement. In developing a regulation or standard, the clarity of the requirement and therefore its likelihood of being complied with have not been given full weight in assessing its economic efficiency or effectiveness.

Compliance promotion activities are designed to tap the motiviations postulated by cooperative theories of enforcement which in turn rely upon the perceived reasonablementss of a regulation. One tradeoff that is often made is between the stringency of a requirement and long term reliability of the control approaches upon which the requirement is based. For example, 80% control may be 99% reliable whereas 95% control may be 80% reliable. Given an understandable reluctance by industry to invest in new and unproven technology, from an enforcement point of view it makes sense to favor the more reliable option and rely upon accepted industry practice Nevertheless, many of our statutes seek the best available control approaches, seeking to force technology and enhanced reliability in use. This is not to say that the technology forcing strategy is necessarily less efficient or effective in the long run, but rather that high rates of compliance will take a longer time to achieve and that projected costs and benefits should take this need for greater enforcement into account.

Other ways in which the design of the regulatory approach can affect compliance include the effect on compliance of individualized permit or other requirements versus general rules of applicability. Rules of general applicability can in theory be easier to communicate to the regulated community, can be easier for inspectors to master, and can offer more standardized enforcement responses. Nevertheless, general rules are historically less economically efficient, and may need to be interpreted as to how they apply to individual facilities; where they do not make sense in individual situations it makes enforcement more difficult. Most important, individually tailored requirements make the facilities or sources far more aware of requirements and the permit specifically interprets how they apply. Therefore, although more complicated to monitor, and more costly administratively, tailored permits — again in theory — can lead to higher levels of voluntary compliance.

Another aspect of regulatory design that will affect compliance monitoring and enforcement is the type of standard. Regulatory requirements generally fall into one of several groups: ambient standards, mass rate or output standards, technology standards, work practice standards, and information requirements. Clear, simple requirements tailored to a source are most amenable to compliance by the source, monitoring by governmental officials and enforcement. Ambient standards, concentrations most related to health and welfare benefits, must be translated to enforceable requirements that pertain to an individual source. This can be accomplished through technology requirements, or performance requirements. Technology requirements are far simpler to enforce and understand from a compliance standpoint: however, they are viewed as economically inefficient since they do not allow more cost-effective substitutions. Performance standards, while more economically efficient, are enforceable only to the extent that the technology exists to reliably monitor performance.

The point behind all of this is that concerns for compliance must begin with the design of a requirement. There are tradeoffs to be made in the simplicity of a regulatory scheme versus the efficiency of that scheme. Our economic analysis techniques that provide answers on efficient and effective regulation, need to take into account the institutional costs of realizing the benefits of the regulation through enforcement and the likelihood of source compliance given the regulatory design.

3. Monitoring Compliance

Compliance monitoring encompasses all those activities undertaken by the regulated community or government officials to collect information on and assess the compliance status of regulated sources of environmental pollution. There are two principles or purposes underlying compliance monitoring strategies used by environmental regulators. The first is that self-awareness and self-monitoring will lead generators of pollution to take essential preventative and corrective action to maintain compliance. The second is that a credible likelihood of detection by government regulators is an essential prerequisite to deterrence. The two principles are linked in practice. Two other purposes served by compliance monitoring are to provide the evidence needed to support enforcement actions for identified violations and to provide reliable statistics on the progress in implementing environmental requirements.

Source self-monitoring and inspections are the most important approaches to monitoring compliance. Ambient monitoring and aerial surveillance are used, but rarely. In theory, compliance monitoring should be a statistically valid indicator of compliance; the methods should be the same as those on which the standard was based; and the methods should be reliable and cost-effective.

In practice, source self-monitoring is not as widespread as regulators would like. First the development of cost-effective monitoring devices has lagged behind regulatory developments. Source monitoring is used extensively in the water NPDES program for all dischargers and for groundwater protection from hazardous waste disposal and storage tanks, but to a far less extent in the air program, where continuous emission monitors are expensive and until recently, were not reliable. At the federal level, the Paperwork Reduction Act seeks to limit information requests of industry and States. Often it is compliance monitoring requirements and other related reports that are sacrificed at the expense of information needed to develop sound regulatory proposals.

Most significantly, no agency can afford to conduct unlimited inspections. The question is therefore one of priorities and the allocation of the scarce inspector resources. To date, priority schemes for inspections are very unsophisticated. They are more focused on one element of the program, the need for breadth of coverage, than they are on targeting inspections on those sources and violation types most likely to yield the greatest benefit from enforcement action. Environmental inspection programs for air and water usually call for inspections of the major sources, generally defined by size and potential environmental impact, at least once per year, biennially for minor sources. In the hazardous waste program the focus has been on land disposal facilities. States and many EPA Regions complain that, given limited resources, meeting national requirements for minimum inspection frequency has prevented them from visiting individual sources that are more likely to be having problems or which locally pose a greater risk.

Recently, there has been some effort using applied statistical techniques to assess how inspection resources can be used more efficiently and effectively . In response to a GAO report and inquiries from Congress, the air program has been developing some alternative models for directing inspections.9/ This application of statistical techniques was based upon some interesting theoretical reasearch at Johns Hopkins University which uses operations research game theory to establish ideal inspection strategies.10/ In addition, in a not yet published study, Duke and Northwestern University economists have tried to correlate federal and state inspections with reductions in effluent in 75 pulp and paper plants over several years.11/ They did identify a statistically significant reduction in effluent after a 4-6 month lag, with little recidivism, following inspections during the period 1977-1985. In the view of this author, further analysis of this data would be desirable, especially on the effects of different types of inspection, the effect of subsequent enforcement action, and the effects of alternative internal management approaches/corporate culture on source compliance behavior.

A result missing from all of the analysis done to date, is a sense of the value of compliance monitoring information. If government officials are constrained from imposing an undue burden on the regulated community in monitoring their ongoing compliance, the question is what is that information worth and who should pay for it? Economic theory would seem to support a source paying for compliance monitoring as the operating cost of doing business and cost of running the pollution control devices that were the subject of significant source investment but regulators need better ways to build these costs and benefits into the initial justifications for and contents of proposed requirements.

What is less easy to address from the perspective of environmental economics is the trend in the most recent set of regulations, from primary use of performance standards and technology to work practice requirements and a growing concern for the regulation of the whole process stream. In what circumstances does it make sense for enforcers to inspect interrelated processes and environmental impacts to ensure the most efficient outcome from the regulatory scheme? Given that work practices are often related to general management, are there economic indicators that correlate well with a plant's environmental performance?

Finally, an issue related to behavior theory and compliance monitoring is whether source self-monitoring should be reported on an exceptions basis or whether there should be complete reporting of all data. Some argue for economic efficiency; only information absolutely needed by regulators should be required. They also argue that reams of data are unuseable by agency officials and that they are better off with less data and more information. Others argue, on behavioral grounds, that regulators can control the quality of data and there is more management attention paid to routine reporting of all monitoring data results.

4. Responding to Violations: Timely and Appropriate Enforcement Response

The concept that enforcement response should be timely and appropriate to the violation is now a key component of current enforcement theory and practice. Swift and sure response is one of the three elements that traditional deterrence theory would deem essential to success. However the concept was only first introduced operationally at the federal and state levels in 1984 through the Policy Framework for State/EPA Enforcement Agreements and program-specific implementing guidance.12/ The Policy Framework was the product of a Steering Committee of state and federal officials from all EPA programs charged with defining expectations, roles, and relationships for an effective national enforcement program.

The Policy Framework defines the timeliness part of the concept in terms of specific points in time at which there should be an initial response to a violation, formal enforcement action (i.e. when informal means are not effective in returning the violator to compliance after a specified period of time), and timely follow through and escalation in the event the violator fails to comply, until full physical compliance has been achieved.

The Policy Framework also defines "appropriate" enforcement response as having three elements. First, there is the appropriate level of formality of enforcement response. An initial violation can be addressed through a full range of informal and formal enforcement tools such as a phone call, site visit, warning letter, notice of violation, formal administrative complaint, and/or proposed law suit. The Policy Framework generally holds that any and all approaches that the government official believes will be most cost-effective are acceptable — unless and until— compliance problems extend beyond a specified period of time and have not been resolved, i.e., it holds that at a certain point in time the response should be formal. The Policy Framework defines formal action as a law suit or formal administrative response, i.e., it is independently enforceable, defines the violation, defines the required response and a date certain for achieving full physical compliance. Another exception is the need for court-imposed action where a violator's schedule to comply exceeds a statutory deadline.

A second element of "appropriate" enforcement response is that it should correct the violation. A third element is that it must include a penalty or other sanction as appropriate to create the necessary deterrence for future violations by that source or other sources. Because only certain enforcement mechanisms can be used to impose a penalty or sanction, where they are needed it requires the more costly formal enforcement responses to be used.

The Policy Framework sets a priority on first meeting timely and appropriate enforcement response for Significant Non-Compliers. Recognizing that requiring timely and formal enforcement action for all violators would overburden limited resources, the Policy Framework only encourages federal and state officials to meet it for non-SNC violations. Furthermore, the concept is used to establish appropriate state and federal roles in delegated or approved States. Under most statutes EPA retains parallel enforcement authority following delegation or approval of a state program. According to the Policy Framework, if the State is pursuing timely and appropriate enforcement response, EPA will defer to the State. However, once the timeframes are passed, EPA will take action if, after discussions with the State, it determines that the State is not moving expeditiously on a case.

The timely and appropriate enforcement response system is built on the concept that it is the willingness of government officials to follow through on less costly enforcement responses and to escalate responses in a timely manner that gives weight and force to lesser responses. Each higher order enforcement response carries with it a multiplier effect in its deterrent value. Initially, to build credibility, officials may be forced to utilize more costly formal administrative or judicial action, but the expectation is that in most instances a simple notice will send violators scrambling to quickly resolve a compliance problem or cooperatively negotiate its resolution.

How has this policy worked in practice? The EPA program offices responsible for implementation of air, hazardous waste, and water programs performed an analysis on the 1986 Fiscal Year experience in a study coordinated by the Office of Enforcement and Compliance Monitoring's Compliance Policy and Planning staff.13/ This analysis will be repeated annually. The analysis found that the concept that enforcement should be "timely and appropriate" is widely accepted as an important measure of the effectiveness of the enforcement effort, but that EPA and the States are wide of the mark in meeting its goals.

Recognizing that there are differences among the programs in the definition of the target universe of Significant Non-Compliers, the timeframes the programs established as goals, and the enforcement tools available, NPDES was able to meet its timely and appropriate criteria about 75% of the time, RCRA was able to meet theirs about 46% of the time and the air program only 22% of the time. The RCRA target for formal enforcement response was 135 days for high priority violations, the NPDES goal is within two quarters of the SNC violation, and the air program is only 120 days. Normalizing for differences in the required timeliness of enforcement response among programs, it appears that all three programs were able to return about 75% of the Significant Non-Compliers into compliance within 270 days.

Some of the reasons for failing to take timely or appropriate enforcement response include inadequate resources, cumbersome enforcement procedures — particularly a lack of simple administrative penalty authorities — and/or a reluctance to pursue formal enforcement action.

At this juncture, there is a real need for research on what should constitute timely and appropriate enforcement response. First, we do not have a good feel for the impact or cost-effectiveness of the various forms of enforcement response. This analysis is complicated because the cost-effectiveness of lesser responses is dependent on the use of more costly and onerous responses. In a dynamic system, how much escalation to the more costly forms of enforcement such as law suits or administrative orders do we have to pursue to underwrite less costly responses? Second, we should carefully examine the value of timeliness of response versus severity of response. There have been some interesting applications of decision theory and economic models in Superfund settlement cases. These cases involve huge sums of money. The models address whether the government should continue to negotiate, sue, or settle, given the strength of the case, its value and needed deterrence.14/

Further, our legal systems for formal response are complex, in part to safeguard personal property and freedoms from unwarranted government intrusion. Nevertheless, there is room for reform. How much is it worth to society to reduce and streamline these procedures, and at what cost to society in terms of guarantees of certain rights? How much should we be willing to invest in more administrative law judges and the like to speed the processing of appeals on administrative complaints and thereby speed the environmental results? These are all questions that are amenable to economic analysis.

5. Imposing Civil Penalties and other Sanctions

Civil penalties and other sanctions such as criminal conviction, shutdown of operations, sewer bans, etc., play an important role in enforcement actions. The imposition of a sanction is a critical third element in deterrence theory. In the past, many enforcement actions merely set forth tailored compliance agreements detailing remedies and schedules for correcting the violation. However, it is now generally recognized that if there is no consequence to violating an environmental requirement (except having to meet with government officials to agree to do what was required in the first place), there is little incentive to undertake any costs of compliance before getting caught. This has proven to be true even when it is broadly understood that clean-up costs will increase substantially if violations are not corrected early and where there has been an actual cost savings from compliance activities.

The Policy Framework for State/EPA Enforcement Agreements addresses three aspects of current operating theory related to civil penalties and other sanctions. First, as noted above, each environmental program is to identify where a penalty or sanction is essential for an enforcement response to be effective, recognizing that penalties cannot easily be sought in each and every case. It is more costly to bring an enforcement action which seeks some sanction both in terms of agency time and resources. This cost is in large part due to the legal protections noted above. The complexity of our enforcement procedures are proportionate to the potential severity of the sanction. The cost in time and resources in seeking penalties also reflects the fact that penalties are more hotly disputed by violators than the fact of the violation and/or needed remedies. Finally, requiring sanctions in all cases of a certain type emphasizes general rather than specific deterrence. In some instances the equities of an individual case could argue against a penalty. Current policy would identify certain violations where sanctions are a must, placing more emphasis on looking beyond the good faith efforts of an individual violator in order to set an example and create an incentive for others to do all they can to comply and avoid a penalty. In such cases, if a State enforcement action contains no penalty or sanction, EPA will seek to pursue its own penalty case to fulfill this need for consistency and broad deterrence on a national basis.

The second aspect of penalties that is addressed in the Policy Framework is the level of a civil penalty. Economic theory has been most directly used in enforcement in this area. The theory, embodied in the Clean Air Act°s Section 120 Administrative Penalty authority and in EPA penalty policies, is that the recovery of the economic benefit of noncompliance is essential to deter would-be violators from seeking economic gain by deferring the required expenditures on environmental pollution controls. The theory has much appeal. EPA first adopted it formally as a minimum desired level of penalty in 1984./15 (Note that the penalty policies not only call for recoupment of the economic benefit of noncompliance, but also would have penalty levels add a gravity component reflecting the severity of the violation, its potential harm, the compliance history of the violator, etc.) Several States also have adopted penalty policies which articulate this philosophy. 16/ However, many States do not share the view that it is necessary to recover the economic benefit of noncompliance to deter violations. Indeed most States simply invoke the maximum statutorily permissible penalty level and proceed to negotiate from there. It is EPA policy that State enforcement actions should attempt to recover economic benefit at a miniumum, but this is not required. However, EPA will seek to take its own enforcement action to recover additional penalties if a state penalty is grossly deficient under the circumstances. The criteria for this determination include whether the penalty bore any reasonable relationship to the seriousness of the violation, and/or economic benefit gained by the violator.

The issue of penalty or sanction is perhaps the most sensitive issue in the state/federal enforcement relationship. If compliance can be achieved more expeditiously by a violator who is willing to agree to a schedule and remedial actions but who will dispute proposed penalties or sanctions, many States would prefer to forego penalties or to agree to a lesser penalty amount than would EPA. This ignores questions of equity, fairness, and effectiveness in the context of general deterrence created by a penalty.

Third, the Policy Framework explicitly recognizes that there are non-monetary sanctions which can have a deterrent effect that may be more powerful than monetary penalties, and that those will be acceptable substitutes. The Policy calls for national guidance as to what alternative sanctions would be acceptable for this purpose. This includes sewer bans, pipeline severance, permit revocation, and incarceration, all of which impose some economic cost.

In relating theory to practice, addressing first the issue of whether a penalty or other sanction is sought, EPA seeks a sanction or penalty in a high percentage of its formal enforcement responses where this authority exists, and the percentage has been increasing. In the RCRA and air programs, EPA sought penalties in 93% and 81%, respectively, of the formal enforcement actions taken. In contrast, the States seem to seek penalties in far fewer enforcement actions. The States sought penalties in 49% of the RCRA enforcement actions and 62% of the air actions for which penalties presumably should have been sought.

These discrepancies reflect the continued philosophical differences between EPA and States on the use of penalties in the overall compliance program. This is particularly true in the municipal compliance arena for water discharge requirements. States have not fully bought into the position that enforcement should be undertaken regardless of the availability of construction grant subsidies for building municipal treatment plants. While many States are indeed seeking penalties as is EPA, a legitimate question remains whether municipal penalties have an effect comparable to the deterrent effect of penalties imposed on industry given the nature of municipal finance. Nevertheless, officials have reported that municipalities seem much more willing to enter into negotiations and settle disputes on violations now that there is a record of high penalty assessments against municipal non-compliers. Several questions remain to be answered: Where do monetary penalties make a difference? How is that difference in behavior manifest? Would other sanctions be more effective?

The second area of actual practice involves EPA's success in recovering the economic benefit of noncompliance as advocated by theory. Based upon a recent study completed by the Compliance Policy and Planning Staff at EPA, it is doubtful that, until recently, penalties levied in many environmental enforcement cases actually approached the levels necesary to recover the economic benefit of noncompliance.17/ However, the 1984 EPA Penalty Policy strengthened the provisions setting recovery of economic benefit as a floor where the concept is applicable. In the 1985 Fiscal Year alone, EPA imposed one third of all penalties imposed in its entire 10 year history and levels were more like what one would expect from the economic models. Thus, despite policy encouraging EPA staff working on enforcement case settlements to recover, at a minimum, the economic benefit of noncompliance, only the most recent penalties approach those levels. Realistically, some of the huge sums that emerge from the very successful EPA model used to compute economic benefit cannot be sustained in all settlement negotiations, although the Courts have upheld EPA policies and citizen groups have used it very effectively in some significant court cases. 18/

Despite the historical failure to fully recover the economic benefit of noncompliance, air and water programs have "reportedly" achieved fairly high rates of compliance.19/ This would suggest that there are other factors at work besides a simple economic calculus on the part of the regulated community. The evidence is that something else is driving compliance besides a simple economic decision. This is particularly significant given that the expected cost according to the economic equation (the probability of a penalty times the expected penalty amount) can be quite small due to the low likelihood of discovering a violation.

The fact that experience is not explained by the most clear application of economic theory to environmental compliance gives one pause, although the theory is clearly one that makes sense as something to strive for. What is not clear however, is what other factors are motivating the behavior of non-compliers and how they can be addressed.

Unfortunately, there is no empirical data on the systematic use of sanctions beyond those associated with monetary penalties. In response to the emerging body of literature that suggests a more sophisticated approach to compliance behavior and the application of economic theory, the Environmental Law Institute, under a grant from EPA's Economic Benefits staff, is currently reviewing the range of these other sanctions available to and used by EPA and the States.

Criminal sanctions are viewed by many as the most effective deterrent in the environmental enforcement arsenal, particularly by those who favor the behavioral models of compliance. Indeed, environmental crimes have gained substantial public support as compared to other crimes.20/ Criminal sanctions are increasingly being sought by federal EPA, the Department of Justice and the Federal Bureau of Investigations and in a growing number of state programs. It is however, generally valid only for intentional circumventing of the law, for negligence under the Clean Water Act, and in all instances of unpermitted dumping into our waterways under the Safe Rivers and Harbors Act. Criminal cases are quite costly and involve very complex procedures. A relatively small but growing number of cases where jail terms have been meted out have begun to change some corporate management ethics. 21/ Pronounced policy on compliance with environmental laws and general compliance programs is no defense (nor mitigating factor for sentencing purposes) for corporate officials charged with the criminal wrongdoing of their employees. Courts will only consider environmental programs with strong oversight and follow through for the specific activities in question. This seems to be having a significant effect on prevention of violations through internal compliance systems and employee compliance incentives within a firm.

Moreover, EPA is seeking to tap these internal corporate incentives by introducing environmental audit provisions into consent decree negotiations in cases where there is a clear pattern of environmental management problems or a pattern of a given type of violation within a company.22/ This creative sanction is also supported by the behavior theory schools which would advocate establishing, to the extent possible, self-policing and internal control systems within a company, tied to environmental compliance. This policy has been supplemented by a campaign by EPA to illustrate by case example, the benefits to industry of sound environmental management and periodic environmental auditing. 23/

EPA is also making increasing use of its Contractor Listing authority whereby a listed company/facility is deprived of the right to be awarded federal government contracts as long as it is on the list. This is mandatory for successful criminal cases and is discretionary for facilities in violation of Clean Air Act and Clean Water Act requirements. The sanction has offered significant economic leverage in several difficult compliance cases. In addition, EPA is placing increasing emphasis on the use of publicity surrounding its enforcement actions and in creative settlements requiring violators to use publicity to enhance deterrence.

6. Clarifying the State/Federal Relationship

It is very important that enforcement be viewed as firm, effective, and fair on a national basis. In 1984 EPA and the States drew up a Policy Framework for implementing State/EPA Enforcement Agreements which set forth clear roles and responsibilities in enforcement. The Policy clarifies the expectations for good performance in implementing a strong enforcement effort, and establishes protocols for advance notification and consultation on all inspection and enforcement matters. In addition it establishes the importance of consistent national reporting of key indicators to assess how effectively the national compliance and enforcement program is being carried out.

A key principle was the introduction of timely and appropriate enforcement response criteria and the definition of what is required in the form of a sanction. If a State is not getting what it should through its enforcement program, EPA will. The ideal of presenting a unified face to the regulated community, and consistency in response has been widely regarded by both industry and state and federal officials alike, as essential steps toward a more effective enforcement program.

Recent evaluations show that these Agreements are beginning to have an effect. However, there are still philosophical differences between EPA and some States which will have to be worked out over time. A key factor here for environmental economists is the fact that while EPA does less than 10-30% of inspections under the delegable programs, it is responsible for about 30% of all formal enforcement actions nationally. Any assessment of improved efficiencies and effectiveness in enforcement response must take into account the role of state and local governments involved in the process.

7. Evaluating Performance and Accountability

Formal reporting and accountability systems that have been established for all environmental enforcement programs set forth five key indicators of performance: 1) the rate of compliance, 2) the progress in addressing Significant Non-Compliers and returning them to compliance, 3) the number of inspections (or number or percentage of facilities inspected), 4) the number of administrative actions taken, and 5) the number of civil and criminal judicial enforcement actions taken. In addition to these indicators, at the federal level, each year efforts are made to assess federal penalty practices and the question of whether enforcement by both federal EPA and States has been timely and appropriate. Furthermore, EPA internally assesses the extent of compliance with consent decrees and any follow up action taken.

There is increasing attention to issues such as penalty collection and state penalty assessment practices or other sanctions but there is no systematic oversight or data collection at this time.

The traditional economic literature would have us examine only the likelihood of having a violation detected and the amount of penalties imposed in reviewing the performance of our enforcement programs. The author would like to see more realistic measures of program effectiveness developed and applied.

V. THE CHALLENGE TO ENVIRONMENTAL ECONOMISTS

Compliance with and enforcement of our environmental laws should be a central issue for environmental economists, but the economics of enforcement has received little attention to date. The economic literature has made a great contribution to the issue of compliance penalties, with the theory that enforcement programs should recover the economic benefit of noncompliance. The theory is widely regarded as a basic tenet of enforcement programs at the federal and state level. Nevertheless, theory does not always equal practice. While it is worthwhile to base our penalty assessment practices upon this theory, we must recognize that compliance motivation and the realities of an enforcement program are far more complex than this simple theory adequately addresses.

The challenge to environmental economists is to provide analysis of empirical data on compliance and to develop decision tools and techniques that are practical and can be used by state and federal officials to make this increasingly difficult task more cost-effective and efficient — starting with the design of regulations to compliance monitoring and promotion strategies to enforcement response and accompanying sanctions. The issues are well defined, the answers are not.

There are several difficulties environmental economists will face in meeting this challenge. The first is the fact that data is not conveniently collected in one data system, nor is it systematically compiled over time. The recent report on Federal Civil Penalty Practices was a painstaking effort for EPA staff, who labored through hundreds of manual calculations in developing data sets on penalty trends and statistics. In the future, this information will be computerized, but at the time of the study it was not. When it is computerized, the data on the environmental effects of a facility, inspection activities, self-monitoring reports, administrative and judicial enforcement actions are often located in different computer systems and they are difficult to cross reference.

The second, and perhaps greater difficulty is the fact that compliance results cannot be explained by economic theory alone. The challenge here is to draw from other disciplines and empirical studies of decision-making within the regulated community to derive a more complete understanding of compliance behavior.

In conclusion, environmental economic research and applications in enforcement and compliance monitoring have been limited to date, which is surprising given the potential payoffs in enforcement and compliance for learning something more about the economic efficiency of our environmental regulations and standards. What information is needed by environmental regulators or by the regulated community on compliance and enforcement that environmental economists can deliver? We need more information on:

- Compliance motivation:

  What motivates compliance? under what circumstances? What is the effect of the form of the requirement i.e. performance, technology or work practice standard, permit or general regulation? the cost and availability of technology?

  What is the effect of corporate culture and management systems design on compliance?

  What is the effect on environmental compliance of the competitive environment/of reputation, in different economic markets?

  Are firms which are financially unsound less likely to comply? What are the implications for targeting of enforcement give the likely difficulties in collecting large penalties from marginal firms which may not be able to pay?

- Compliance Practices:

  What are the quantifiable benefits of specific environmental management practices? What are the costs to regulated entities of past, present and future noncompliance?

  How can liability costs be incorporated into internal economic decision making? What algorithms can be used by company managers (financial managers, engineers and plant managers) to factor future environmental liabilities (both compliance related and risk related) into internal decision making? How can activities which have short term costs (environmental auditing, waste minimization) be better presented to show long term benefits?

- Compliance Monitoring:

  What is the value of information on compliance in different settings given the risks to the environment and public health from a violation and its probability of occuring?

  Who should pay for compliance monitoring? What are the implications for types of inspections? for required compliance self-monitoring and reporting in regulations and permits? for research and development on compliance monitoring techniques?

  How frequently and what kind of inspection should be undertaken, at what facilities, to maximize deterrence and provide accurate compliance statistics within a given budget?

- Enforcement Response:

  Given the environmental risks from noncompliance, how much should governments spend to ensure high levels of compliance?

  What are the multiplier effects and cost-effectiveness of individual enforcement response options, given their interdependence? in different contexts?

- Sanctions:

  How effective are alternative sanctions in providing the necessary disincentive to non-compliance, in what settings? Is the sanction equitable, and fair in addition to being effective?

  What are the effects on competition and markets of uneven enforcement? How consistent does national enforcement need to be?

  What levels of penalty should be imposed for reporting and record keeping violations? (The economic theory of recovering the costs of non-compliance does not work well here.)

Do high penalties and enforcement actions targeted at a few influential actors within the regulated community have a greater deterrent impact than lesser penalties imposed on a larger percentage of violators? How does either approach compare in terms of equity, cost-effectiveness, etc.?

- Tax Policies, Technical Assistance and Subsidies:

  What are the effects of tax policies and capital markets on compliance? What are the compliance implications of alternative policies?

FOOTNOTES

1/ *See* "Project 1 Report on Economic Efficiency of Enforcement and Enforcement Related Monitoring (March 1986) Unpublished Draft Report to the EPA Office of Policy Planning and Evaluation by Professors Wesley A. Magat at Duke University, W. Kip Viscusi at Northwestern University, Richard Feckhauser, Harvard University and Peter Schuck at Yale University (March, 1986).

See also "Enforcement of the Clean Water Act: Theory, Policy and Practice". Unpublished Draft Report prepared by ICF, Incorporated for the Office of Policy, Planning and Evaluation (January 31, 1987).

2/ *See* " Study of Literature Concerning the Roles of Penalties in Regulatory Enforcement", September 1985, Thomas Charlton, Compliance Policy and Planning Branch, Office of Enforcement and Compliance Monitoring.

3/ *See* John T. Scholtz, "Cooperation, Deterrence, and the Ecology of Regulatory Enforcement", Law and Society Review, Volume 18, No. 2, 1984.

4/ *See* Robert A. Kagan and John T. Scholtz, "The °Criminology of the Corporation° and Regulatory Enforcement Strategies", Enforcing Regulation, edited by Keith Hawkins, and John M. Thomas, Law in Social Context Series, 1984, Kluwer-Nijhoff Publishing.

See also "Psychological Factors Influencing Compliance", February 7, 1985, by Dale T. Miller, Study for the Federal Statutes Compliance Project, Department of Justice, Ottawa.

5/ *See* note 2 *infra*.

See "Analysis of Environmental Compliance Theories", June, 1987. Unpublished draft report submitted to EPA°s Office of Policy Planning and Evaluation,° by Errol Meidinger, Barry Boyer and John Thomas of SUNY at Buffalo.

See also "Environmental Compliance and Corporate Culture: Methods and Motivations", April, 1987. Draft workplan submitted to EPA°s Office of Policy Planning and Evaluation, by Kurt Fischer at Tufts University, Center for Environmental Management.

7/ *See* Note 4 *infra*.

8/ *See* Memorandum from Thomas L. Adams Jr., titled Summaries of FY 1988 Enforcement Priorities", (April 16, 1987)

9/ *See* "Final Phase II Report Stationary Source Inspection Modelling Project: Allocation of a State°s Inspection Resources, July, 1987, by Barnes Johnson, EPA°s Office of Policy Planning and Evaluation.

10/ *See* "Mathematical Techniques for Optimization of Enforcement Strategies" (August, 1985), Jerry A. Filar, Assistant Professor, Department of Mathematical Studies, The Johns Hopkins University.

11/ See "Project 2 Report on Economic Efficiency of Enforcement and Enforcement Related Monitoring" March 30, 1987. Unpublished draft report to EPA's Office of Policy, Planning and Evaluation, Professors Wesley A. Magat at Duke University, W. Kip Viscusi, at Northwestern University.

12/ See Memorandum form A. James Barnes, Deputy Administrator, titled "Revised Policy Framework for Implementing State/EPA Enforcement Agreements" (August 26, 1986)

13/ See Memorandum from Thomas L. Adams, Jr. Assistant Administrator for Enforcement and Compliance Monitoring, titled "Report on the Implementation of the Timely and Appropriate Enforcement Response Criteria" (January 15, 1987)

14/ See Memorandum from Lloyd Guerci, Director CERCLA Enforcement Division, titled "Release of Updated Computer Model for Decision Tree Analysis of Proposed Superfund Settlements" (June 3, 1987). This model was developed in early cooperative work by EPA staffer Bob Fuhrman, then of the Office of Policy, Planning and Evaluation and Brad Wright of the Office of Waste Programs Enforcement.

15/ See Policy on Civil Penalties, EPA General Enforcement Policy #GM - 21 (February 16, 1984).

16/ See Environmental Law Institute, "State Civil Penalty Authorities and Policies" (September 30, 1986) report prepared for the U.S. EPA, Compliance Policy and Planning Branch, Office of Enforcement and Compliance Monitoring.

17/ See Memorandum from Thomas L. Adams, Jr., titled "Report on Federal Penalty Practices" (to be issued), Carol Hudson Jones, Compliance Policy and Planning, Office of Enforcement and Compliance Monitoring.

18/ See Cheseapeake Bay Foundation v. Gwaltney of Smithfield, Ltd., 611 F. Supp. 1542 (E.D. Va. 1985), Aff'd 791 F. 2d 304 (4th Cir. 1986) Cert. pending on unrelated issue, S. Ct. No. 86-473.

19/ See "Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies, United States: A National Review" Cheryl Wasserman, October, 16, 1984 prepared on behalf of the Organization for Economic Co-operation and Development. Group of Economic Experts.

20/ See United States Department of Justice, Bureau of Justice Statistics Bulletin, January 1984, in which 60,000 people were asked in a public opinion poll, to rank the severity of particular crimes. Environmental crimes ranked seventh after murder but ahead of heroin smuggling and skyjacking.

21/ Since 1986, federal judges imposed over 144 years in prison for environmental crimes, of which over 36 years will be served and over $4.6 million in fines were levied. About 70% of those convicted have been individuals, mostly senior corporate managers.

22/ See "Environmental Auditing Policy Statement; Notice" Federal Register Vol. 51, No. 131, Wednesday, July 9, 1986, page 25004. See also Memorandum from Thomas L. Adams Jr., titled "Final EPA Policy on the Inclusion of Environmental Auditing Provisions in Enforcement Settlements". (November, 14, 1986)

23/ See Benefits of Environmental Auditing, Case Examples, December, 1984, Arthur D. Little, Inc., prepared for EPA's Regulatory Reform Staff.

Benefits to Industry of Environmental Auditing, August, 1983, Arthur D. Little, Inc., prepared for EPA's Regulatory Reform Staff.

Annotated Bibliography on Environmental Auditing, Sixth Edition November 1986, EPA's Regulatory Reform Staff.

Annotated Bibliography on Environmental Management, First Edition, November 1985, Arthur D. Little, prepared for EPAs Regulatory Reform Staff.

ENFORCEMENT ISSUES IN SELECTING POLICY INSTRUMENTS

Lee G. Anderson
University of Delaware

August 1987

Paper prepared for the AERE Workshop on Enforcement in Environmental and Natural Resource Regulation

INTRODUCTION

In the literature on both pollution control and fisheries management, there are many comparative studies of the relative effectiveness of policy instruments such as taxes, quotas, input restrictions, etc. Most give particular emphasis to the ability of the instrument to achieve economic efficiency. Little attention, however, has been given to comparisons according to the relative ease of enforcement. Enforcement is defined to include monitoring activities to identify rule breakers, and prosecution activities wherein the formal guilt or innocence is ascertained and penalties are assessed and implemented where guilt is proved. This is a serious deficiency because if there are practical constraints to enforcement, what may appear to be the overall superior instrument, may be a dismal failure as far as actually achieving program benefits.

The purpose of this paper is to identify enforcement issues that can be important in practical policy application and, where appropriate, to evaluate standard policy instruments on their ability to face problems caused by these issues. At the outset, however, it should be noted that there are no clear winners and losers as there are in evaluations based on economic efficiency of producing effort. In fact, in many cases, it is not possible to compare general classes such as standards vs. taxes. The necessary comparison must be between certain types of standards and certain types of taxes. Due to the author's background and predilection, the presentation will be primarily in terms of fisheries management.

Some of the discussion can be directly applied to pollution control programs but for a complete description of some of the specifics, a separate analysis will be required.

The first section will briefly review the problem of environmental and natural resource regulation when enforcement issues are considered. The second will describe the current monitoring and enforcement options and policies under existing fisheries management law. This will provide a better background to understand the overall enforcement problem. The third section will list and describe in detail some of the important enforcement issues and how they can affect the bottom line results of various regulation programs. Many of the issues may not be obvious to those who concentrate on the more traditional aspects of environmental and resource policy.

The new information provided in this paper, when added to the existing literature, will provide a framework for choosing the best instrument or instruments for different situations each with its own economic, environmental, and political peculiarities. In certain situations, this more complete framework may result in a different choice of management programs than would be the case if enforcement issues were ignored.

THE ENFORCEMENT OF ENVIRONMENTAL AND NATURAL RESOURCE PROGRAMS

The prime economic justification for pollution control and fisheries regulation is market failure. Because of externalities and/or open-access to the relevant resource, the unregulated economy will not maximize the value of goods and services produced. Optimal regulation programs are based on the premise that they can reallocate inputs associated with these processes such that the net present value of goods and services produced over time will increase.

While the above may be an overly simple statement of the economic foundation of environmental policy, it does capture its essence. Most practitioners would stress, however, that a regulation program can not be justified unless the present value of gains is greater than any implementation and enforcement costs. Until very recently, however, these costs have been (explicitly or implicitly) treated as fixed annual amounts that are merely subtracted from the gross gains. Recent work has shown, however, that the problem is much more complicated. [Downing and Watson, 1974; Harford, 1978; Hucke, 1978; Downing and Kimball, 1980; Storey and McCabe, 1980; Beavis and Walker, 1981; Brady and Bower, 1982; Downing and Kimball, 1982; Krupnick, Magat and Harrington, 1982; Richardson, 1982; Ullmann, 1982; Lee, 1984; Linder and McBride,, 1984; Malik, 1984; Martin, 1984; Sutinen and Andersen, 1985; Tietenburg, 1985; Anderson and Lee, 1986; Milliman, 1986; Russell, Harrington, and Vaughn, 1986; and Beavis and Dobbs, 1987]. To set the stage for the analysis to follow, it will prove useful to review some of this discussion.

The net gains from a fishery regulation program can be represented by the schematic in Figure 1 where all items are measured in present value terms. The important point to note is that there are many types of enforcement related costs, and these costs will vary with the type of policy instrument used and to the degree to which bureaucracies are willing or able to pursue their enforcement activities. That is, while the net benefit of an unregulated fishery may be the difference between the value of the output and the cost of fishing effort, the net benefits of a regulated fishery must take into account all costs which follow as a direct result of the regulation program.[1]

Compliance costs are those initial expenses born by firms as they prepare to change their behavior to come into compliance with regulations. With provisions such as gear restrictions, this can involve the purchase of new gear or the adaptation of old. Taxes, on the other hand, may require no capital expenditures per se, but only the acquisition of the appropriate forms and the knowledge of how to prepare them.

Items such as lobbying costs to amend regulations or obtain variances and litigation expenses to get regulations overturned or at least have their implementation delayed or to obtain exemptions can also be classified as compliance costs. Rather than meet specified standards or behavioral patterns, the firm attempts to get the rules changes so that current operations will be in compliance.

---

[1] Similarly to measure the gains from a pollution program, it is necessary to look beyond the difference between the net value of marketable output and abatement costs. All costs directly attributable to the implementation and enforcement of the regulation program must also be considered.

Avoidance costs are the expenses firms may undertake to hide non-compliance or to make it appear as if they are in compliance with the rules. For example, boats could have two sets of nets on board if there are mesh size regulations, or boats could off load part of the catch at out of the way ports or to other vessels if there were trip limit regulations.

While the first two costs are imposed on regulated firms, the other two classifications are government expenses. Initial government implementation costs are the start-up costs of getting the program operating. If the regulation is similar to other government programs (i.e. quotas for a newly managed fishery when quotas are used in many other fisheries) and existing organizations or agencies can perform the task, these costs could range from little more than preparing memos to assign tasks to existing staff to hiring and training new personnel. If the regulation is radically different, however, these costs can be quite high as new agencies or branches are established, personnel are hired, standard operating procedures are established, etc.

Government enforcement costs include both monitoring and prosecution activities. The purpose of the former is to encourage compliance and to detect deviance. The latter can range from formal court proceedings to administrative hearings and all of the negotiations, bargaining, and other transactions that accompany them. See Sutinen (1986) and Wasserman (no date).

While considering these extra costs when measuring the net value of a regulated fishery, it is necessary to determine how each varies both with the type of instrument used, and the level at which it is used. While it will be beyond the scope of this paper to present a formal model of this optimization process, sufficiently precise policy conclusions for the analysis to follow can be derived from a rather simple statement of the problem. For more detail see Anderson and Lee (1986).

In the most general case, the problem for the management agency is to select the appropriate combination of policy instruments and to allocate implementation and enforcement inputs such that the net gains from the regulated fishery are maximized. To solve this problem it is necessary to know how the types of instruments used and levels to which they are applied (i.e., the type and size of the total quota or the nature and the extent of trip limits) and the inputs used to support their application (i.e., the number of person years of labor allocated to the various aspects of enforcement) will affect each of the elements in the net gains equation.

At the risk of being redundant, it is important to emphasize that it is not a matter of maximizing the net benefits from the fishery per se, or of minimizing regulation costs. It is the algebraic sum of all terms in the equation that is crucial. Further, each one can be affected by the types and extent of regulation. The value of fishery output will be affected according to the efficacy of the regulation program in actually changing industry participants' behavior (i.e., how much it will actually reduce effort in an overextended fishery.) In addition the quality of the marketable product may be affected for good or ill. For example, a regulation plan which would increase the length of the Pacific Halibut season would allow for more fish being sold fresh. On the other hand, in other fisheries, a restrictive quota which can easily be circumvented by landing at out of the way ports and trucking it to central marketing areas may increase the average time it takes to get the fish to the processor or the consumer and hence will likely result in a decrease in quality.

The type of regulation can also affect the cost of producing effort. This has been the focus of most of the economic analysis of policy instruments. The universal conclusion is that traditional measures such as gear restrictions, closed seasons, closed areas, and total quotas cause fishing effort to be produced at a higher cost than is necessary. Further, except for the latter, they will not be completely efficacious in the long run as industry participants modify their fixed and variable inputs to maximize profits subject to the constraints posed by the regulations. On the other hand controlled access type regulations such as taxes, individual transferable quotas, and to some extent, license restriction programs produce incentives for efficient production. See Rettig and Gitner, 1978; Pearse, 1979, Sturgess and Meany, 1982; Beddington and Rettig, 1984; and Anderson, 1987.

That the type of regulation program can affect avoidance and government implementation and enforcement costs should be obvious from the discussion above. It follows therefore, that in order to provide a complete economic evaluation and comparison of policy instruments, it will be necessary to give more detail to how each will affect the five elements in the net gains equation.

To go one step further, however, the problem facing most resource agencies is more difficult than a straightforward application of a simple maximization problem which allows them to optimally choose policy instruments and to allocate implementation and enforcement inputs. In the first place, they seldom know the exact nature of the gains function to be maximized. In the second place, there may will be other (explicit or implicit) maximands (mostly dealing with income distribution and other politically sensitive issues) that are often the driving force behind agency operations. Incidently, the agencies may be just as uncertain about the true nature of these other maximands as they are about the net gains equation. Finally, they are constrained by an operational budget. In reality then, the agency faces a constrained maximization problem and one where there is a good deal of uncertainty about the nature of the maximand.

Because of the constraint, the allocation of the agency inputs is just that much more critical. For this reason, the relative, as well as the absolute effects of the various instruments on items in the net gains equations will be of interest. Further, the budget allocation problem is made more difficult because it is not just a matter of knowing how to allocate funds among different regulation inputs per se, but also between regulation inputs as a group and research activity devoted to obtaining a better picture of the biological, ecological, and economic aspects of the world in which they operate. That is, they have to use scarce inputs to better define their maximand and the institutional and ecological constraints under which they operate. In this regard, comparison of policy instruments with respect to what is currently known or can be known about their effects will be important.

In summary, when comparing policy instruments, it is necessary to go beyond the efficiency effects on the production of effort (or the efficiency effect on the production of waste product abatement in the pollution case). The absolute and relative effects on industry compliance and avoidance costs as well as government implementation and enforcement costs can also be very important in determining the proper regulation program.

MONITORING AND PROSECUTION PROCEDURES IN CURRENT FISHERIES POLICY

The purpose of this section is to briefly describe existing enforcement procedures for fisheries policy in the United States today. Knowledge of the strengths and limitations of the current operational institutions is necessary for complete policy evaluation.

Monitoring

There are two basic monitoring modes: at sea and dockside. According to the National Marine Fisheries Service Office of Enforcement (Pallozzi and Springer, 1985), the former includes (1) observers placed on board fishing vessels, (2) boarding and inspecting vessels at sea, (3) ship patrols to observe fleet activity by remaining at sea for long periods of time, (4) boat patrols to observe fleet activity by vessels which have more limited range and endurance due to inability to handle rough weather and to remain at sea for long periods of time, and (5) aircraft patrols to observe fleet behavior from the air.

Dockside modes include (1) monitoring landings, gear types, record keeping, etc. at the point where the harvest is discharged, (2) inspection of dealers or processors at steps further down the production process as the fish is passed to on to final consumers, and (3) investigations such a reviewing self reporting forms prepared by fishermen but also including undercover operations, radio monitoring, data analysis, use of informants, etc.

Each monitoring type has different relative abilities to detect deviance from specific types of regulation, and just as important, each has different relative costs. In general at sea monitoring is more expensive but it does offer the potential to observe all aspects of the fishing operation. The trade-off is that while dock-side monitoring is relatively less expensive it is also remote from the fishing activity. The key question is whether this remoteness will frustrate the purpose of management by allowing activities to take place between harvesting and landing which make it more difficult to determine if the fishing operation was conducted in a legal manner.

There is obviously more to this question than a simple trade-off between cost and relative efficacy. First, at a more technical level, there is a serious question of what is the most important regulation objective and how is the best way to enforce it. For example, it may be that in a certain fishery both total catch limits and prohibitions on retaining small individuals appear to have potential value as possible regulations.[2] If the total limit is thought to be more important, and since it can be adequately enforced at dockside, it makes sense to regulate the catch of young fish in a way that can also be enforced dockside. For example, in this hypothetical case, mesh size restrictions might be better, in an overall sense, than incidental catch restrictions even though the latter may have an absolute advantage in reducing mortality of small fish _if_ enforced with an effective on-board monitoring program. While the mesh size restrictions may not be as effective as a strictly enforced incidental catch limit, the marginal cost of enforcing it in combination with the dockside monitoring of total catch, may be so much lower than enforcing the incidental catch limit at sea, that the overall net benefits of the mesh size restrictions will be much higher. All else equal then, it may make sense to select the combinations of regulations such that only one monitoring mode is required even if some of the governing instruments chosen are second best in an absolute sense. (Pallozzi and Springer, 1985.)

Along these same lines, an optimal management program may produce distributional or biological effects, that, when viewed in isolation, appear sub-optimal. For example, consider a fishery which requires restrictions on the number of small fish taken.

---

[2] The purpose of the total limit is to protect the spawning stock while the constraint of landing small individuals is to allow them to reach larger size before harvest.

This can be accomplished by minimum mesh size limits or area restrictions if fish migrate over their life cycle or if the stock breaks into patches of relatively old or young fish in various parts of the ocean at different times of the year. However, because of the relatively high cost of enforcing at sea, it is possible that simple possession prohibitions of small mesh nets, or of small fish, which can be enforced at dockside, may be the regulations which produce the highest net gains.

Possession restrictions on small mesh nets are not only enforceable dockside but they may be the only practical way to assure compliance. If more than one mesh size were allowed on board, at sea monitoring would be required to ensure that the smaller mesh nets are not used. And even with at sea monitoring, it may not be possible to achieve compliance if successful prosecution requires the boarding officer to find the illegal net in the water. (See detailed discussions below).

While mesh size restrictions with prohibition on possession of small mesh nets can be enforced at dockside, they can cause distributional problems. Some boats may make a habit of switching from fishery to fishery in the course of a trip according to fish availability, relative prices, etc. Therefore they will desire to carry more than one set of nets (each legal for at least one fishery) in order to operate as profitably as possible. So while a regulation that only allows one set of nets on board at any one time, may significantly reduce overall enforcement costs and indeed may be the only way that is really enforceable, such a plan may significantly increase costs to all or part of the fleet. In essence while the regulation costs to the government will fall significantly, there will be some large increases in regulation costs to the industry. An angry constituent complaining to his or her legislator that a "silly bureaucratic law" requires the boat to return to shore and unload catch before changing nets is likely to find a sympathetic ear.

There are ways to overcome this problem, however. For example, as part of its controlled access program where fishermen are given quasi-property rights in the fisheries, Australia has instituted a user pays system where the owners must pay a levy to cover some of the operational costs of enforcement. This levy is currently around 38% of the assessed enforcement costs, and there are plans to increase it to 50%. Under this system industry pressure for supplementary regulations which must be enforced at sea have significantly declined. (Lilburn, 1986, p. 155)

As a second example of a management program with side effects which appear irrational, consider prohibitions on the possession of small fish. While such restrictions would be easy to enforce on land and while this would encourage individuals to operate in areas or times when small fish are less frequently caught and also to shift locations if small fish are taken on the first set, they will result in the dumping of any small fish that are taken whether or not they are dead or alive when returned to the sea. In some instances however, the loss of this product may be less the savings in enforcement costs as compared to other types of regulation.

However, complaints focusing exclusively on the discard of small edible fish can make a fisheries agency look very bad.

As another point in favor of dockside enforcement, it should be noted that under current law considerable leeway is granted to dockside monitoring. If used effectively with appropriate regulations, this flexibility could make dockside monitoring a very powerful monitoring mode. For example, section 1857(1)(G) of the Magnuson Fishery Conservation and Management Act (MFCMA) makes it illegal to "ship, transport, offer for sale, sell, purchase, import, export, or have custody, control, or possession of, any fish taken or retained in violation of this Act" or its implementing regulations, permits, or Governing International Fishing Agreements. The same section imposes strict liability on fish processors or merchants in that violations do not require elements of willfulness, intent, or even knowledge. (Jacobson, et al., 1987, p. 112). "Sting" operations on distributors, could seriously damage the market for illegal fish and hence the incentive for harvesters to ignore fishing restrictions.

Prosecution Procedures Under Existing Laws[3]

The first step in the prosecution procedure is to have a violation documented by a Coast Guard or a NMFS agent. The agent then prepares an Offense Investigation Report (OIR) which names the violator, describes the offence and the conditions surrounding the arrest, and discusses any pertinent details regarding the particular case. This report is forwarded to the regional law enforcement division of NMFS where a lawyer determines if the case is legally sound. If so, the attorney issues a Notice of Violation and Assessment (NOVA). The contains the same information in the OIR but it also assesses a penalty and names the party responsible for paying it. The penalty can be assessed, either jointly or individually, on the boat owner and the captain to discourage the former from giving orders for deviant behavior or from hiring individuals who might engage it in and to directly discourage the latter from disobeying regulations.

The maximum size of the penalty is determined by law ($25,000 for each violation), but the size of the assessment depends upon the seriousness of the infraction, the number of offenses by the respondent, ability to pay, and the perceived effect of the size of the fine as a deterrent to others and upon the incentives of the respondent to contest the decision. The latter two elements are especially important with respect to overall program success and the optimal allocation of the limited amount of attorney time.

---

[3] The material for this section is drawn from Dolan and Kundin (1985), Frailey (1985), Matera (1985), Nies (1985), Sutinen (1986), and Sutinen and Hennessey (1985). The summary provided here will be all too brief due to space constraints, but the reader interested in the actual prosecution of fisheries regulation will find all of the above very interesting. Similarly those interested in the enforcement of pollution control laws, should obtain at least as comprehensive a treatment of enforcement procedures in that area as is provided for fisheries in these papers.

The respondent can agree to pay the fine, attempt to negotiate a reduced fine, or request a hearing. The last two responses are not mutually exclusive. If negotiations between the attorney and the respondent fail, the case can be brought before an administrative law judge who, after reviewing all the evidence and hearing arguments from both sides, can dismiss the case, or change the amount of the fine in either direction. If the respondent disagrees with this decision, he or she can appeal to the administrator of National Oceanic and Atmospheric Administration, the agency in which the National Marine Fisheries Service is located. If the outcome is still not satisfactory, the respondent has the opportunity to move to the federal district court system and appeal the case as far up as the Supreme Court.

At the end of the negotiation and/or appeal procedure, a final order is issued with the statement of the fine. If the assessment is not paid in a timely manner, the Attorney General is authorized to recover the amount in federal district court. Since the Attorney General has a full range of duties to perform, the final collection of the fine is sometimes only undertaken with encouragement and pressure from the NMFS regional attorney's staff. Any failures to actually collect assessed fines in a timely way obviously reduces incentives to comply with regulations. In fact even the successful postponement of fines can significantly lower their present value and, as such, can diminish their retarding effect on illegal behavior. This is especially true given the likely high discount rate of fishermen.

In certain instances the regional attorneys have another penalty option besides fines. The MFCMA allows councils to require participants to obtain a fishing permit as a condition of working in Federally regulated fisheries although not all management plans take advantage of this option. Where permits are required, the attorney may revoke the permit of the respondent as part of the penalty, the length of the revocation time being subject to the same set of conditions that determine the amount of a fine. In addition, the permit may be revoked for those individuals who have outstanding unpaid fines. Therefore, enforcement agents working in fisheries where permits are required have another tool at their disposal. Given the potential loss of income that could result from a permit revocation, it is a powerful tool indeed.

The MFCMA also authorizes enforcement officers to seize a fishing vessel including gear and other items on board that reasonably appear to have been used in violation of the act. Independently, or as part of a vessel seizure, officers may also seize fish illegally taken and retained. Although this would seem to have a very strong deterrent effect, it is only used for very serious or repeated violations. (Jacobson, et al., 1987, p. 116.) Exceptions are the surf clam and scallop fisheries, where seizures are common for catches violating a minimum size restriction.

Technically under the law, fishermen can be charged with criminal violations if they interfere with enforcement activities. This is includes such things as resisting arrest, threatening enforcement offices etc.. However, there appears to be a move to decriminalize fishing violations as much as possible, and so this section of law is used very sparingly. For example a Japanese vessel was alleged to have fished in US waters without a permit, refused the admittance of a Coast Guard boarding party, attempted to evade seizure, and positioned itself to ram the arresting Coast Guard vessel. Even so the decision was made to only press for the civil penalty of forfeiture of the vessel and its catch.

ENFORCEMENT ISSUES IN SELECTING POLICY INSTRUMENTS

Introduction

The purpose of this section will be to introduce issues that can be of importance in selecting fishery management regulations. In particular, the focus will be on the effect each can have in the determination of the size of the elements in the net gains equation depicted in Figure 1. The discussion of each issue will be brief, but the goal will be explain exactly how the issue is related to the direct or related costs or the effectiveness of regulation. The issues will be discussed in no particular order.

Ease of Governmental Implementation

There are two separate points here. The first is the legality of various types of regulations. Obviously if certain types of

NET GAINS FROM REGULATED FISHERY

VALUE OF FISHERIES OUTPUT

MINUS

COST OF FISHING EFFORT

MINUS

COMPLIANCE COSTS

(COST OF INDUSTRY ADAPTING OR
COMPLYING WITH REGULATIONS)

MINUS

AVOIDANCE COSTS

(INDUSTRY COSTS TO AVOID
DETECTION OR TO FACILITATE
NON-COMPLIANCE)

MINUS

INITIAL GOVERNMENT IMPLEMENTATION COSTS

MINUS

GOVERNMENT ENFORCEMENT COSTS
INCLUDING MONITORING
AND PROSECUTION COSTS

FIGURE 1

controls are forbidden by legislation or precluded by constitutional guarantees, they will not be enforceable. However, constitutionality is often not determined until after legal challenge. For example, state fisheries laws which prohibited or unduly restricted residents of other states from fishing were declared unconstitutional on the basis of equal protection under the law.

The issue is more complicated, however. For one thing, in addition to the legality of the regulation, there must also be suitable procedures to enforce it. As an extreme example, although there is a per vessel quota in the Nova Scotia herring fishery, there is no regulation stating that the fishermen must weigh the fish they sell. Therefore they can't be charged with misreporting their catch because they can declare ignorance of the exact amount landed. (Peacock and MacFarlane, 1986, p. 226.)

Further, authority over fisheries is often vested in many jurisdictions including the various states and the management councils. What is permitted in one state may be illegal in another. At worst, this can cause inter-jurisdictional conflicts and at best it can cause time-wasting delays in co-operative management as individual states pass conforming legislation. This has been a serious problem in enforcing the Common Fisheries Policy of the EEC; the overall agreement included some stipulations that were forbidden under the laws of some of the member countries. Those countries couldn't enforce the rules until the domestic laws were changed and when the others saw this, they were hesitant about enforcing them if it would disadvantage citizens. (EEC, 1986)

Besides legality, ease of implementation also depends upon how enforcement will merge with the existing institutional structure. As discussed above, the more a new program differs from existing rules, the more difficult and more costly it will be to implement it. In this regard, traditional regulations such as total quotas, gear restrictions, closed area or seasons (programs which cause inefficiency in the production of effort) will be easier and less costly to implement, at least on the margin. On the other hand, an individual transferable quota program (the management device that seems best able to encourage efficiency) will be quite expensive to set up. There will be the cost of determining the initial distribution of the quotas. Because of legal constraints, it will be necessary to set up an appeal procedure for those participants who feel they were treated unfairly. Experience has shown that because of the high potential rewards, a large proportion of those who are potentially eligible will appeal, either to get a quota or to increase their share. (Lowman, 1986.) There is also the cost of instituting a control mechanism to keep track of ownership and transfers of quota shares and of matching an individual's catch with its his or her purchased or rented quota for a given period. The high implementation costs, however, can be moderated by low continuing enforcement costs. New Zealand, for example, expects that the present value of the reduction of all at-sea monitoring and much of the dockside efforts that were required with more traditional management will more than compensate for the high initial implementation costs of their ITQ system. (Carouthers, 1987.)

Period at Risk When in Non-compliance

The issue here is the length of time the regulatee is at risk of being identified as being not in compliance. Obviously, all else equal, the longer the period, the more effective the regulation will be. In fisheries, area closures are at one extreme. The fisherman is only at risk during the period in which he or she is in the closed area. In most cases, once a fish is landed, it is impossible to determine where it was landed. The other extreme would be an individual transferable quota program where a specific annual catch limit is given to specific individuals or firms. In the first instance, the boat could make a dash into a closed area when fishing is expected to be extremely productive, and would be safe again if it could just get to the open area before being detected. On the other hand, if total catch records can be crossed checked with dock agent reports, or better yet, with company income tax forms, deviance from the annual individual limit could be detected as long as the records remain unaltered and available.

Other regulations fall in between. Gear restrictions such as mesh size limits have a very short period of risk because the vessel must be caught with the net in the water. Prohibitions on possession of certain types of fish have medium periods of risk because infractions can be detected from the time of harvest to at least the time of discharge and sometimes longer if adequate "paper tracks" on the transfer of sale through the markets are available. For example, the harvesters of "short lobsters" are often identified by tracing back from where illegals are found on the market. Possession prohibitions on gear can have a period of risk of the length of the trip for short term capital such as type of net, to years for long term capital such as engine horsepower or vessel displacement.

Ease and Cost With Which Industry Participants Can Achieve Ability to Comply

There are at least two separate points to be considered in industry compliance. First, if compliance requires a significant change in operating behavior, the costs in terms of obtaining the requisite capital equipment or acquiring the human capital can be very high. Obviously, the higher the compliance cost, the lower will be the net benefits of the regulated fishery even when there is complete acceptance of the program. At the same time however, higher costs will also encourage deliberate non-compliance or using avoidance activities to conceal non-compliance. While avoidance activities will decrease industry compliance costs, they will result in an even lower net present value of benefits from the regulated fishery because of the excess pressure on the stocks.

A separate point is the speed with which the required changes can reasonably be accomplished and hence the speed with which industry behavior is actually changed. If one can make a reasonable argument that it takes a long time to obtain the physical or human capital to easily comply on a day to day basis, detections of deviant behavior may prove of little use, because the firm can persuade the court or the administrative law judge that all reasonable steps have been taken to come into compliance, but full adherence to the regulation will simply take time.

At the surface it may appear simple to solve this problem. All else equal, choose that procedure which has the lowest cost of industry compliance. However, all else is not always equal. Not only are there other efficacy and efficiency effects associated with each regulation, but the same regulation will often affect heterogeneous industry participants differently. Therefore it can be quite difficult to select one with the lowest overall compliance costs. The measurement of compliance costs and the selection process will be made more difficult because each segment of the industry will propose regulations that puts the brunt of costs on their competitors.

Most fisheries regulations have compliance costs, but it is difficult to make any general rankings. Gear restrictions will obviously have costs but their size depends upon the exact type of restriction, and the divergence between it and the normal operating mode. Closed seasons and closed areas can have compliance costs if fishing in other areas or seasons is permitted and economically rational (the firm has the choice to cease fishing) and it cost more to operate under the restrictions. Likewise possession limits can have higher costs due to sorting and to search costs to locate appropriate sized fish.

The compliance costs of ITQs will be mostly transactions costs. It will be necessary to work with the agencies responsible for issuing, maintaining, and transferring the quotas and there may be added costs of landing the fish so that it can be checked against a quota. There will also be the transaction costs of buying, selling, renting, or leasing of ITQs such that the full economic efficiencies of the program can be achieved. While these costs may be quite high initially, it is likely they will decrease as formal and informal networks are developed.

The purchase of ITQs from the government or from other participants is a transfer payment and not a compliance cost in a social sense. However, from the private perspective of an individual operator they will be considered as such and the market price of ITQs will be a factor in the decision to comply with the program. This statement should be interpreted with care, however. A high value of fishery will have high ITQ prices, but there will be incentives to cheat when the returns are high no matter what type of regulation is used.

Ease of Distinction Between Honest Mistakes, Sloppy Practices, and Deliberate Cheating

To use an example, the problem here is to differentiate between fishermen who do not comply with say, size restrictions, because: (1) They had difficulty ascertaining average individual size given the total amount of fish landed, the available measuring equipment, and the ease of working it on an open deck in rough seas; (2) The catch was undersized because inadequate training or supervision of workers or because they simply failed to do their jobs on a given day; or (3) The boat deliberately fished for small individuals, or at least failed to discard that part of the catch that was known to be illegal. When it is difficult to distinguish between actions, it is difficult to know what to do on a case by case basis. Accordingly the firms will attempt to have their actions judged as honest mistakes, or at worst, a one time unintentional error.

Prosecuting the first and third types of activities with equal fervor has certain moral ramifications. In addition it can decrease popular support for the program which will make overall enforcement that much more difficult. The issue of equal treatment of the second and third types is similar but not so clear cut. The problems can be reduced by having the severity of the penalty depend upon number of previous or similar violations. This must be done with care, however, because if the initial penalty is too low, the first violation may be viewed as a "freebie" and everyone will be encouraged to take it.

Long term economic efficiency may require enforcement agencies to, initially at least, work with firms which have the first two types of problems to help them come into compliance. On the other hand, immediate and sure punitive actions which cause expected losses to be greater than gains from cheating are necessary to eliminate the third type of behavior.

It is difficult to rank regulation types as to how they handle this issue, but some tentative conclusions are possible. After the initial communication of regulation's stipulations has occurred, it is fairly easy to distinguish deliberate cheaters on quota regulations (both ITQs and total quotas) and closed seasons. These appear to be fairly black and white issues. Once the total quota is captured or if you don't have an ITQ, you should not be fishing. Similarly fishing during the closed season cannot be called a mistake. Area closures are fairly straightforward especially if an individual is caught in the middle of a 100 square mile closed area with all his or her navigational equipment working. The issue is not so clear cut when the areas are small or the individual is caught only slightly over the border. The seriousness of the last problem can sometimes be reduced by instituting buffer zones around the actual area meant to be closed. Gear restrictions can be troublesome unless they are simple prohibitions or otherwise unambiguous. Possession restrictions by species or size are straightforward at least as long as the fish remain in round form. However, when there is a mixed catch, it can sometimes be difficult for the fisherman to sort out the catch and land only permitted individuals.

Initial vs. Continued Compliance

This is a particularly important issue in pollution control because regulation almost always involves a change in operating procedures requiring new or modified capital equipment. Therefore, program cannot be successful unless firms come into initial compliance by obtaining the appropriate equipment. While it is relatively easy to detect initial compliance, having the equipment is not always enough to guarantee a reduction in effluent. In most instances, proper maintenance and operation of the pollution control equipment is also required. Nonetheless, the relatively low cost of initial compliance monitoring and the increased probability of continued compliance it produces normally, makes it highly productive.

This is an important issue for gear restriction fisheries regulations as well because initial compliance often corresponds to the procurement of a piece of capital equipment. However with regulations such as fishery wide total quotas, area closures, the distinction between initial and continued compliance is not so clear. Therefore, with most fisheries regulations there is no analogous one time check on initial compliance. Except for permanent gear restrictions and perhaps some safety devices, fisheries enforcement will have to be continuous to be effective.

In those cases where the distinction does apply, say, a limit on length, horse power, or displacement, monitoring costs will be quite low. However, their overall effectiveness may leave something to be desired. Due to other margins that operators can use to increase effort, these limits place an elastic control on effort and cause effort to be produced in efficiently.

Ease With Which Requirements Can be Communicated

Obviously individuals cannot conform to regulations, even if they desire to co-operate, if they can not understand exactly what is expected of them. There are at least two parts to this problem. First there is the simplicity or complexity of the regulation. A simple prohibition of a particular type of gear is quite easy to understand. But with prohibitions on harvesting certain species in certain areas with certain gear types which change over the course of the year (as was the case in the first cod, haddock, and yellowtail flounder plan prepared by the New England Fishery Management Council in 1977), it is very difficult to know what is permissible at any point in time.

Second, there is the language or framework that is used in posing the requirements. It is important to use generally recognized measurement and analytic methods. For example, if there is to be a trip limit, it should be defined in the way the particular fishery measures its output. For example surf clams are measured in cages, groundfish in tons, and salmon by the number of fish.

The way monitoring is performed can reduce thee problems if the agents view part of their role as showing industry participants exactly what the rules mean in terms of their every day behavior and how they can change their operations to come into compliance.

Overall management costs can be reduced if the agents can produce the ability as well as the motivation to conform. (Wasserman, no date).

Ease With Which Non-compliance Can Be Disguised

This issue is self-explanatory but there are some subtle distinctions which should be illuminated. By use of detection avoidance activities, fishermen are able to appear to be in compliance. Avoidance activities make it more difficult for agencies to detect the actual amount of fishing. They can be anything from underreporting to subterfuges such as fishing or landing fish at night or the use of remote ports or fishing grounds. If the avoidance activities are costless (i.e., 150 pounds of fish are put in a standard box when the rule of thumb used by enforcement officers to measure total catch is 125 pounds per box) then the problem is only one dimensional. Non-compliance will cause total catch to be higher than the desired amount. However if avoidance activities are relatively expensive (i.e., carrying two sets of nets or landing fish in secondary ports), then the problem is two dimensional. The management objectives will not be achieved and in addition, extra economic resources will be drawn into the fishery lowering the net economic gains.

Where avoidance costs are low, all else equal, non-compliance will normally be quite high because the gains from cheating will be higher than the costs. This will be especially true if the individual can see little private gain from compliance and if others are not complying either. However, when avoidance costs are relatively high, the amount of non-compliance will decrease.

Accordingly, when studying the possibility of deliberate avoidance that is likely to accompany any particular regulation, both the types and the costs of possible avoidance activities, and the potential gains from non-compliance must be considered. In high value fisheries, incentives for avoidance will be there, and unless the regulatory program can be designed accordingly, one or both dimensions of this problem may result.

Ease With Which Agents Can Detect Non-Compliance Such That It is Admissable as Evidence

The key to this issue is the last few words. Detecting non-compliance in such a way that the agent is very sure that it exists is one thing, but obtaining evidence that will stand up in court is quite difficult. As Perry Mason would say "Does the evidence show beyond a reasonable doubt that the individual is in non-compliance."

There are at least three points here. One has to do with the nature of the regulation. For example, it is not as easy to obtain admissable evidence on prohibitions of the use of small mesh nets as it is on one prohibitions a small mesh net. From the time a Coast Guard boat is spotted on the horizon until it can board a fishing vessel, there is often plenty of time to switch from a small mesh net to a large mesh net. Unless the boarding team can find the small mesh net in the water, evidence of use in problematic. Even if the small mesh net is found to be wet and to contain fish, it is not possible to argue beyond a reasonable doubt that spray did not get it wet and the fish from the large mesh net fell into it. On the other hand, with prohibitions on small mesh nets, it is fairly easy to establish that a small mesh net is on board.

A second point is ease with which agents can be trained to obtain appropriate evidence. This is probably more of a problem in pollution regulation than in fisheries, but it is only a matter of degree. The technical issues discussed in the previous paragraph may require some legal training as to what types of evidence are formally admissable, but that is not the real point. When the regulations are in terms of parts per thousand emitted on average over a two week period for pollution control or average number of individuals per pound of product landed for a several thousand pound catch for regulation of a scallop or clam fishery, there can be some very difficult technical and statistical measurement problems. See Russell, Harrington, and Vaughn (1986). If it is very difficult to train individuals to draw their samples correctly and to use the equipment, and also to retain their services once they're competent, then enforcement costs will be high and enforcement efficacy will vary over time.

Finally it is easier to obtain convincing evidence if scarce monitoring resources can be focused on those individual who are more inclined to non-compliance. Therefore those regulations which have an easy "trigger" to identify cheaters, even if that particular information will not stand up in court, can be relatively more effective, all else equal, because monitoring resources can be used more carefully.

Degree to Which Personal or Social Benefits from Compliance Can Be Demonstrated

The private decision to comply with a regulation is very complicated. Certainly the relative private benefits and costs of compliance play a part. But in addition, the individual's view of himself as part of a society and the responsibility that such membership entails is also important. Therefore, the more that fishermen can see individual benefits from the program in general and from their cooperation in particular, the more likely they will comply. Along the same line, programs which are perceived as providing benefits to the fishery as a unit, or perhaps even for the society as a whole, will be easier to enforce on those individuals who consider these things to be important.

Regulation is required in the first place because private actions lead to non-optimal social results, therefore it follows that private compliance with a properly formulated control program will also involve net costs. Nonetheless, the higher the perceived benefits, the higher will be the compliance rate. More correctly, individuals will suffer no guilt pains for non-compliance with a program that has no perceived benefits to anyone.

In some ways this argument is compounding. For example, regulations which are thought to be unenforceable will likely be viewed as producing no benefits, therefore the motivation to comply will be low.

Potential for Citizen Co-operation in Identifying Offenders

The degree to which participants or others are willing and able to provide information on the non-compliance of others (with or without a reward) will obviously lower monitoring costs which should increase overall net benefits. The motivation to help monitor is obviously important. This is partly related to the personal benefits participants or the general public expect to obtain from successful program operation. In commercial fishing, such gains are possible if two or more user groups are competing for a stock, and informing on rival groups will improve the relative position of one's own group.

The ability to identify offenders is also important here and this will related to the nature of the regulation. Casual observers will be able to discern non-compliance with area and seasonal closures much easier than with gear restriction or catch limits.

Likelihood of Encouraging Rentseeking Behavior by Industry and of Administrators Being Susceptible to It

The more that individual participants feel that pressure applied on the management agency will yield favorable results, the more they will engage in such rent seeking behavior. While rent seeking can sometimes provide at least temporary gains for private individuals, they produce net losses at a social level.

There is an extensive literature on rent seeking with respect to government operations in general, but the logic applies to fisheries management as well (see Bhagwati, 1982; Buchanan, Tollison, and Tullock, 1980). Management institutions which directly or indirectly encourages industry input will motivate resource waste because all industry participants will be inclined to participate in such activities in order to remain competitive.

There is a delicate balance here, however. Industry can often provide information that is extremely useful to the management process. But at the same time, it is not hard to find examples where industry participation must be viewed as lobbying rather than a public service. Those management institutions which are most sheltered from industry pressure will not encourage wasteful behavior, but at the same time, they may miss the opportunity to obtain important information at low costs.

Although the type of management institution is important in determining the amount of rent-seeking activity, the type of regulation can also be a factor as well. For example gear restrictions and closed areas seem to invite specific lobbying because they normally hurt various sectors of the fleet differently. Individual interests will be highly motivated to get an exclusion for their gear type, or what amounts to the same thing, a more restrictive control on the gear type of their competitors. Petitions to change the boundary lines for closed areas by a few degree of latitude or longitude for much the same purposes are also common occurrences. Not only do these actions slow up the implementation of management, the loopholes provided can strongly affect the potential for gains. Also the higher the number of variances to a specific ordinance, the harder it is to communicate the rules to industry and the more difficult it is to monitor.

Ease With Which Illegal Activities Can Be Detected Under Various Conditions

Compliance with most regulations will likely vary across time, space, type of vessel or gear, species sought, etc. because the gains and losses of obeying the rules can critically depend on such operational parameters. For example, susceptibility to certain gear types may vary throughout the year and so cheating may be more profitable in January than it is in July. Likewise, the stock may be exploited by two different types of gear or different groups each of which has their own enforcement problems.

It is important, therefore, that the ability to enforce be flexible enough to be able to handle the various situations that are likely to prevail. More important, of course, is the ability to be able to work in those particular times where the motivation for non-compliance is relatively strong and when non-compliance can have the most serious deleterious effects on the objectives of management. For example, Zavolta, Strand, and Swartz (1986) found that 5% of the fishermen accounted for the great majority of the commercial stripped bass harvest in Chesapeake Bay. To be effective a regulation program will have to be enforceable on this individuals even if the other 95% are not affected.

Relative Ability to Which Enforcement is Efficacious With Respect to Different Management Objectives

While economic efficiency is obviously an important management objective by which to compare different types of regulations, it is seldom, if ever, of most concern to real world decision makers who more often focus on employment and distribution of income effects. In some instances, while it may be difficult to enforce a certain type of regulation such that one type of goal is achieved, obtaining sufficient compliance for another goal may be relatively easy. Therefore, rating regulation types is not possible unless the full range of management objectives is considered.

Ease With Which Benefit Based Priorities For Enforcement Can Be Identified

While a management program may prohibit several activities, the net benefits of perfect enforcement of each part may not be the same. For example, with fast growing cohorts, gear restrictions will provide more protection to small fish earlier in the season. Therefore the gains in terms of increased value of fisheries output from enforcement of gear restrictions will vary throughout the season. Since enforcement resources are always scarce, information on when and where to enforce can be quite valuable. Therefore it is important to be able to identify not only the change in behavior that is likely to result from varying degrees of enforcement, but also the net gains the changes in behavior will produce. The need for government to be able to identify these differences is obvious, but it is also important for industry to see them as well so that the rational behind what may otherwise appear as a random enforcement can understood. Without this understanding, a lack of respect for the management program could develop which could seriously reduce the net benefits of management.

SUMMARY

(to be added later)

REFERENCES

Anderson, Lee G. 1987. "Expansion of the Fisheries Management Paradigm to Include Institutional Structure and Function." Transactions of the American Fisheries Society.

Anderson, Lee G., and Dwight R. Lee. 1986. "Optimal Governing Instrument, Operation Level, and Enforcement in Natural Resource Regulation: The Case of the Fishery." American Journal of Agricultural Economics 68(3): 678-690.

Beavis, Brian, and Ian Dobbs. 1987. "Firm Behaviour under Regulatory Control of Stochastic Environmental Wastes by Probabilistic Constraints." Journal of Environmental Economics and Management 14: 112-127.

Beavis, Brian, and M. Walker. (1981). "Random Wastes, Imperfect Monitoring and Environmental Quality Standards." Journal of Public Economics 21: 377-387.

Beddington, John R., and R. Bruce Rettig. 1984. Approaches to the Regulation of Fishing Effort. FAO Technical Paper #243. Rome: AFO. 39 p.

Bhagwati, Jagdish N. 1982. Directly Unproductive, Profit Seeking Activities. Journal of Political Economy 90: 988-1002.

Brady, G. L., and B. T. Bower. 1982. "Effectiveness of the U.S. Regulatory Approach to Air Quality Management: Stationary Sources." Policy Studies Journal 11 (September).

Buchanan, James M., Robert D. Tollison, and Gordon Tullock (eds.). 1980. Toward a Theory of Rent-Seeking Society. Texas A&M University Press, College Station, Texas.

Carouthers, S. 1987. Personal communication.

Dolan, Hugh J., and Cynthia L. Kundin. 1985. "Administrative Law Judges and Proceedings." Workshop on Fisheries Law Enforcement, University of Rhode Island, October 1985, 22 p.

Downing, Paul B., and William D. Watson, Jr. 1974. "The Economics of Enforcing Air Pollution Controls." Journal of Environmental Economics and Management 1: 219-236.

Downing, Paul B., and James N. Kimball. 1980. "Enforcing Administrative Rules: The Case of Water Pollution Control." NBER Conference Paper Series, No. 66 (National Bureau of Economic Research).

Downing, Paul B., and James N. Kimball. 1982. "Enforcing Pollution Control Laws in the U.S." Policy Studies Journal, September, p. 55-64.

European Economic Communities (EEC). 1986. "Report on the Enforcement of the Common Fisheries Policy." Com (86) 301. Brussels. 30 p.

Frailey, Margaret H. 1985. "Problems of Case Management." Workshop on Fisheries Law Enforcement, University of Rhode Island, October 1985, 22 p.

Harford, Jon D. 1978. "Firm Behavior Under Imperfectly Enforceable Pollution Standards and Taxes." Journal of Environmental Economics and Management 5: 26-43.

Hucke, J. 1978. "Bargaining in Regulatory Policy Implementation: The Case of Air and Water Pollution Control." Environmental Policy and Law 4: 109-115.

Jacobson, Jon, Daniel Conner, and Robert Tozer, editors. 1987. "Federal Fisheries Management: A Guidebook to the Magnuson Fishery Conservation and Management Act." Revised edition. University of Oregon Law School. 161 p.

Krupnick, A., W. Magat, and W. Harrington. 1982. "Understanding Regulatory Decision-Making: An Econometric Model." Policy Studies Journal 11(1): 44-54.

Lee, D. R. 1984. "The Economics of Enforcing Pollution Taxation." Journal of Environmental Economics and Management 11: 147-160.

Lilburn, Bruce. 1986. "Management of Australian Fisheries: Broad Developments and Alternative Strategies." *In* Fishery Access Control Programs Worldwide: Proceedings of the Workshop on Management Options for the North Pacific Longline Fisheries. Alaska Sea Grant Report No. 86-4, p. 141-187.

Linder, S. H., and M. E. McBride. 1984. "Enforcement Costs and Regulatory Reform: The Agency and Firm Response." Journal of Environmental Economics and Management 11: 327-346.

Lowman, Dorothy. 1986. "Legal Aspects of Effort Limitation." *In* Limited Access Alternatives for the Pacific Groundfish Industry, edited by Daniel D. Huppert, NMFS, p. 105-118.

Malik, A. A. 1984. *Noncompliance and the Economic Theory of Pollution Control*, unpublished Ph.D. dissertation, Johns Hopkins University.

Martin, Lawrence W. 1984. "The Optimal Magnitude and Enforcement of Evadable Pigovian Charges." Public Finance 39(3): 347-358 (Netherlands).

Matera, Marguerite. 1985. "Regional Management of Fisheries Law Enforcement Cases: The Assessment of a Civil Penalty." Workshop on Fisheries Law Enforcement, University of Rhode Island, October 1985, 11 p.

Milliman, Scott R. 1986. "Optimal Fishery Management in the Presence of Illegal Activity." Journal of Environmental Economics and Management 13(4): 363-381.

Nies, Thomas A. 1985. "Evaluating Enforcement Effectiveness." Workshop on Fisheries Law Enforcement, University of Rhode Island, October 1985, 11 p.

Pallozzi, Morris M., and Steven C. Springer. 1985. Enforcement Costs in Fisheries Management: The Alternatives." 14 p.

Peacock, F. Gregory, and Dougald A. MacFarlane. 1986. "A Review of Quasi-Property Rights in the Herring Purse Seine Fishery of the Scotia-Fundy Region of Canada." In Fishery Access Control Programs Worldwide: Proceedings of the Workshop on Management Options for the North Pacific Longline Fisheries. Alaska Sea Grant Report No. 86-4, p. 215-230.

Pearse, Peter H. 1979. "Symposium on Policies for Economic Rationalization of Commercial Fisheries." Journal of the Fisheries Research Board of Canada 36(7): 711-866.

Rettig, R. Bruce, and Jay J. C. Ginter, editors. 1978. Limited Entry As a Fishery Management Tool. Seattle: University of Washington Press.

Richardson, G. 1982. "Policing Pollution: The Enforcement Process." Policy Studies Journal 11(1): 153-164.

Russell, Clifford S., Winston Harrington, and William J. Vaughn. 1986. Enforcing Pollution Control Laws. Resources for the Future, Washington, D.C.

Storey, D. J., and P. J. McCabe. 1980. "The Criminal Waste Discharger." Scottish Journal of Political Economy 27: 30-40.

Sturgess, N. H., and T. F. Meany, editors. 1982. Policy and Practice in Fisheries Management. Proceedings of the National Fisheries Seminar held in Melbourne, 1980. Canberra: Australian Government Publishing Service.

Sutinen, Jon G. 1986. "Enforcement of the MFCMA: An Economist's Perspective." American Fisheries Society's Annual Meeting, September, 1986, Providence, RI. 32 p.

Sutinen, Jon G., and Peder Andersen. 1985. "The Economics of Fisheries Law Enforcement." Land Economics 61(4): 387-397.

Sutinen, Jon G., and Timothy M. Hennessey. 1985. "Fisheries Law Enforcement: An Overview and Analysis." Workshop on Fisheries Law Enforcement, University of Rhode Island, October 1985, 30 p.

Tietenberg, T. H. 1985. Emissions Trading, An Exercise in Reforming Pollution Policy. Resources for the Future, Washington, D.C.

Ullmann, A. 1982. "The Implementation of Air Pollution Control in Germany." Policy Studies Journal 11(1): 141-152.

Wasserman, Cheryl. No date. "Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies - United States: A National Review." U.S. Environmental Protection Agency.

Zavolta, S. M., Ivar E. Strand, and David G. Swartz. 1987. "Supply Response in the Harvest of Striped Bass." Conference on Economics of Chesapeake Bay Management II. Annapolis, Maryland. 27 p.

SESSION II

Modeling Optimal Enforcement Strategies

Clifford S. Russell, Vanderbilt University
"Game Theory Lessons for Structuring Monitoring and Enforcement Systems"
Clifford S. Russell, Vanderbilt University

Gary V. Johnson and Thomas S. Ulen, University of Illinois
"The Behavior of the Firm When Facing Uncertain Enforcement of Ex Ante Environmental Regulation and Tort Liability Standards"

Arun S. Malik, University of Maryland
"Markets for Pollution Control When Firms are Noncompliant"

GAME THEORY LESSONS FOR STRUCTURING MONITORING AND ENFORCEMENT SYSTEMS

Clifford S. Russell*

Prepared for the Association of
Environmental and Resource Economists'
Workshop On Environmental Monitoring
and Enforcement: Theory and Practice

Newark, Del., 13, 14 August 1987

---

* Thanks are due to the Alfred P. Sloan, Andrew W. Mellon, and National Science Foundations for support of the research on which this paper is based. Many people have commented on earlier papers in this vein, but this paper reflects especially helpful comments from V. Kerry Smith and Mark Cohen of Vanderbilt University. The usual disclaimer applies.

Whenever laws or regulations require individuals or firms to act in ways contrary to their narrow self-interest some provision must be made for monitoring — to see that those subject to the requirements comply with them — and enforcement — to punish noncompliance so as to encourage compliance by the offender and others. Without such provisions, the legal requirements effectively become only calls for voluntary sacrifice. But the provision of monitoring and enforcement requires resources; how much and what kind depending on the requirements at issue. In general it is not worthwhile to devote enough resources to catch every instance of noncompliance, nor is it practical to institute penalties sufficiently draconian to discourage all noncompliance when the probability of being caught is less than one. This much is rough common sense, codified and extended by economists writing on crime and punishment (eg: Becker 1968 and McKean 1980).

At the same time, there is considerable, albeit fragmentary, evidence that in a variety of environmental contexts, current monitoring and enforcements efforts by U.S. EPA and its state agency partners are producing lacklustre results. (For a catalog of this evidence from air, water and hazardous waste management contexts, see Russell 1987). For example, self-monitoring reports of over 770 major water pollution point sources were examined by the U.S. General Accounting Office (GAO) for periods in 1978/79 and 1980/82. These reports revealed that fully one quarter (U.S. GAO 1980, U.S. GAO 1983) of the sources produced significant violations of their discharge permit terms during the periods of examination. ("Significant" was defined to mean a greater than 50 percent violation on one or more pollutant limits continuing for at least four months.) About sixty-five percent of the reports revealed at least one violation of a permit requirement for one month or more. And 30 percent of the sources violated one or more permit terms by some amount for at least six months.

Considering that these data come from self-reporting records for which no independent verification is or ever can be available, it is reasonable to view them as representing lower bounds on actual violation rates.[1] Whether actual violation rates are "too high" is, of course, a difficult question if a technically correct answer is required. That answer would require estimates of the damages to be avoided by producing lower rates along with the costs of the efforts required to produce those lower rates; and implicitly, an understanding of the mechanism by which the greater effort encouraged the improvement in compliance. None of this information and related modeling is easy to produce, but by far the greatest obstacle to finding the "socially correct" amount of monitoring and enforcement effort, and hence the correct violation rate, is the requirement for estimates of damages avoided. While much effort and ingenuity have gone into improving methods and data bases for damage (benefit) estimation in the pollution control field, the available measures are still for the most part crude and aggregate relative to the rather fine-scale marginal damage estimates required for monitoring and enforcement system design.

Therefore one must judge the evidence on compliance and violation on a much looser basis; and on such a basis it is hard to disagree with the theme of the several GAO reports on this subject that the current rates are unsatisfactory and threaten the attainment of ambient quality improvements being sought under the various laws. (See, for example, GAO 1979, 1980, 1983, 1985.)

This is not to say that EPA and the states are unaware of the problem. While the first decade of experience under the major environmental laws of the early 1970s involved concentration on getting control technology in place ("initial compliance"), recently there has been a recognition that improving continuing compliance must become a priority. (Wasserman 1984). But there is every reason to expect rather tight budget limits on any enforcement activity. Certainly this is the message of current general budget stringency. It is also implied by past levels of monitoring effort as revealed in surveys of state agencies. For example, a Resources for the Future survey in the early 1980's found that large air pollution sources were visited by state agencies' enforcement personnel for discharge measurement purposes on average once every 8 months, while large water pollution sources were visited on average once every 5 months (Russell 1982). These rates should be viewed in the context of permit terms that generally involved limits on mass discharges per hour (the most common choice for air permits) and per day (most common for water permits) (Russell 1982). The mean visit rates cited above, if one assumes that each visit results in 8 hours of measurements, imply that about 0.4 to 0.7 percent of the possible events (violation or compliance choices) are being sampled for large sources. (See, for example, Casey, et al. 1983.) Or, said another way, if the visits were random, the probability of being monitored for permit-term compliance would be in the neighborhood of 0.005.[2]

Further, such evidence as has been gathered strongly suggests that these low probabilities of being monitored are not being offset by the levying of very large fines for discovered violations. Thus a survey reported in Russell, Harrington and Vaughan (1986) of state experience shows a range of values of expected penalties per discovered violation from zero to over $2500. In only two of 17 states were fines large enough to bring the expected value above $1000 and in only six was that value even above $100.

The above observations imply that suggestions for improvement in monitoring and enforcement system for pollution control should take account of several limitations:

- they should not depend on the availability of damage estimates
- they should be predicated on continuing budget tightness
- they should not depend on the introduction of extremely large fines to make up for infrequent monitoring and consequent low probability of being caught.

One feature of the existing system will not and should not be taken as given. That is the fact that most air and water pollution monitoring of point sources involves preannounced visits (Russell et al. 1986). Whether or not this completely vitiates the incentive features of the enforcement system is an empirical question tied to features of source controllability: how long it takes to bring a non-controlling source into control and how much it costs. For the following discussion, it is assumed that random, unannounced monitoring visits are possible. That these are the rule in nuclear power plant and hazardous waste handling facilities encourages one to believe that assumption not particularly far fetched (eg: USNRC 1983).

This paper outlines a route to improved monitoring and enforcement design that does reflect the above limitations, a route based on a game theoretic view of the monitoring and enforcement problem. In the process, the important problem of measurement error is addressed and its implications for system design explored. The result is a method that, for the same budget limitation and size of fine, allows the attainment of higher compliance levels than does simple application of random visit timing, where the information on past detection is thrown away after any resulting enforcement action is taken. That is, using past violations to influence probabilities of future monitoring visits provides an extra incentive for compliance. The impact of the incentive varies with other features of the problem, in particular with the importance of the fines being levied and the size of the errors of inference.

The Model For Single Periods

It will be convenient to begin with a particularly simple version of an agency monitoring problem. In it there will be N identical sources of pollution all subject to the same limit on their discharges. Each source can save C dollars per period by not complying with the limit. The agency responsible for enforcing the discharge limits can choose in principle to monitor or not monitor each source during each period. If monitoring is undertaken, it costs M dollars per visit.[3] The agency, however, faces a budget limitation, B, which it will later be convenient to express as a fraction of the cost of monitoring every source in a period: $r = \frac{B}{M \cdot N}$

The monitoring instruments available to the agency are not perfect but rather display random errors. The distributions of these errors are such that, in combination with the agency's rule for deciding on the existence of a violation, they imply the following probabilities:

α = the probability of falsely identifying a violation when the source is in compliance (Type I error)

β = the probability of falsely identifying as in compliance a source that is in fact in violation. (Type II error)

The sources are assumed to be risk neutral in this context; that is, they make decisions for the purposes of the monitoring game on the basis of expected monetary value.

The Game

The essence of the game for a source is that the agency and the source jointly determine the payoff for a particular period by their choices of strategy, the agency choosing whether or not to monitor and the source choosing whether or not to comply. Using the assumptions and notation just developed, a source's payoff matrix is as follows:

		Source Strategy	
		Comply	Do Not Comply (Violate)
Agency Strategy	Monitor	$C + \alpha F$	$(1-\beta)(F+C)$
	Do Not Monitor	C	0

In the game context, with the possibility of random, unannounced monitoring visits, a natural question to ask is: what probability of monitoring, call it m, the agency would have to choose in order to give such a source the incentive to comply in every period, even though it is not monitored in every period? This may be determined by looking at the expected value of the two alternative source strategies when m is the probability of being monitored.

$$E\ (\text{comply}) = m(C+\alpha F)+(1-m)(C)$$

$$E(\text{violate}) = m(1-\beta)(F+C)+(1-m)0$$

keeping in mind that these are costs, and solving for the value of m such that $E\ (\text{comply}) < E\ (\text{violate})$ gives after manipulation that:

$$m > \frac{C}{(1-\beta)(C+F)-\alpha F}$$

is sufficient to provide an incentive for source compliance. In later discussion, it will be convenient to define F as a multiple, f, of C, so that $F = fC$ and the key value of m can be written as:

$$m > \frac{1}{(1-\beta)(1+f)-\alpha f}$$

In table 1 a few illustrative values of f, α, and β and the implied values of m are provided.

Table 1: Monitoring Probability That Induces Compliance As A Function Of Fine And Error Sizes*

Relative fine $f = \frac{F}{C}$	Type I error α	Type II error β	Monitoring probability m
1.2	0.05	0.2	0.585
2.2	0.05	0.2	0.408
3.2	0.05	0.2	0.312
1.2	0.01	0.01	0.460
2.2	0.01	0.01	0.318
3.2	0.01	0.01	0.242

* $m = \frac{1}{(1+f)(1-\beta)-f\alpha}$

Budgets and the Extent of Noncompliance

The monitoring probability just defined, if faced by everyone of N identical sources, would provide an incentive for every source to choose to comply. But actual measurement would find the fraction $m\alpha$ in violation in every period. Expanding the observed fraction by the sample fraction implies, not surprisingly, that $\alpha \times 100$ percent of the sources could be expected to appear out of compliance if all sources were monitored in a given period.

The dollars spent on monitoring the fraction m of the sources would be NmM. Expressing this as a fraction of the cost of complete monitoring gives us: $m = \frac{NmM}{NM}$.

But suppose, as seems reasonable, that the available budget for monitoring would not support this much effort. How are the attainable fraction of sources in compliance and the budget required for that attainment related?

First, observe that the agency cannot proceed in a situation of limited budget $(r < m)$ to announce a monitoring probability r applying to all sources, for then none would have an incentive to comply. Rather it must announce that the probability m applies to a suitably chosen subsect N_r, such that $\frac{N_r \cdot m \cdot M}{N \cdot M} = r$, the available budget.[4] That is $N_r = \frac{N \cdot r}{m}$. Then the expected fraction of sources out of compliance in a period can be inferred in two parts:

- those sources that do not expect to be monitored
- those sources that are monitored and are expected to be found out of compliance.

Denoting the expected fraction of sources in violation as ε, we obtain:

$$\varepsilon = \frac{N - N_r}{N} + \frac{N_r\alpha}{N}$$

$$= \frac{N - \frac{N.r}{M}}{N} + \frac{\frac{N.r.\alpha}{M}}{N}$$

$$= 1 - \frac{r}{M} + \frac{r\alpha}{M}$$

$$\text{or} \quad \varepsilon = 1 - (1-\alpha)\frac{r}{m}$$

Thus, if r is fixed, this ε is the smallest fraction of sources that can be expected to be in violation.

On the other hand, it may be of interest to determine what budget would be required to produce an expected violation fraction equal to some ε. This is given by: $r = \frac{(1-\varepsilon)}{(1-\alpha)}m$

for $1 \geq \varepsilon \geq \alpha > 0$.[6] Note for future reference that $\frac{\partial r}{\partial \varepsilon} = \frac{-m}{1-\alpha} < 0$. That is, the required budget declines as the requirement on the fraction of sources in violation is relaxed.[5] Illustrative values of required budgets implied by assumed f, α and β, are provided in table 2.

Extending The Game To Repeated Periods Or Plays

In the agency-source game described above, the agency takes no account in its future plans of the discovered behavior of any source. Violators are fined, but that is the end of it. Another way of saying the same thing is that the only penalty for a discovered violation is a one-period fine, F. In the tax-compliance literature, models have been developed that make use of information on past behavior (as discovered by monitoring, of course) in defining either the future probability of monitoring or the future fine for a discovered violation or both. For example, Landsberger and Meijison 1982 allow both fine and probability to vary within a two state model, while Greenberg 1984 varies the monitoring probability only, but uses a three-state model.

Table 2: Budget Fractions Required To Obtain Chosen Compliance Levels

Allowed fraction not complying (ε)	Relative fine (f)	Type I error (α)	Required (Single play) budget (r_{sp})
0.2	1.2	0.05	.493
0.2	3.2	0.05	.263
0.1	1.2	0.05	.554
0.1	3.2	0.05	.296
0.2	1.2	0.01	.372
0.2	3.2	0.01	.196
0.1	1.2	0.01	.418
0.1	3.2	0.01	.220

These approaches to the construction of monitoring and enforcement systems have attracted attention because they appear to allow the attainment of given levels of compliance with the regulations in question using lower levels of spending by the authorities. Systems closely akin to repeated games with grouping of regulatees are found in practice as well. For example, the U.S. Internal Revenue Service is widely believed to use data from past audits to define the probability of a current year audit. And new regulations aimed at controlling ground water pollution (Fortuna and Lennett 1987) effectively group landfill operations for current monitoring purposes on the basis of past detection of problems.

A repeated monitoring game may conveniently be thought of as a markov process in which the states are groups among which the regulated parties move according to the results of particular plays of the embedded game. The probabilities of transition depend on the frequency with which the regulated parties are monitored when in each group. Discovery of a violation when in one group results in reassignment of the party to another group with different monitoring probability. In Greenberg's systems there are three groups. Being caught in violation while in group 1 results in reassignment to group 2 and a subsequent violation in group 2 would result in reassignment in perpetuity to group 3, in which auditing is constant (every year for an income tax system) and perpetual. A party monitored and found in compliance in group 2, however, is returned to group 1. Greenberg shows that in the absence of errors in the audit, and when the parties do not discount future costs, all parties in group 2 will comply and no party will ever be assigned to group 3. Monitoring probabilities in the first two groups can be small fractions of the levels necessary to stimulate compliance in a single play game situation. All parties in group 1 will find it optimal to violate because of their second chance in group 2.

There are several ways to interpret such systems. If one views the regulated parties as generally attempting to comply, grouping is simply a way of concentrating resources on those least capable of successfully doing so. An alternative, where compliance choice is taken to be an open question, is to see the variation of monitoring probability on the basis of past record as an integration of the two elements of the expected cost of noncompliance — size of fine and probability of its exaction. In this sense it corresponds to enforcement approaches involving higher fines for repeated offenses.

One may also interpret such extensions of single-play enforcement games as involving the added element of threat: If you don't do as I say you should, I will undertake a course of action neither of us really wants — checking up on you so frequently for such a long time that you will face a compliance "hell" and I will have to spend a considerable sum on your compliance alone. Behave in a reasonable fashion and I will offer occasional opportunities to relax. In the repeated game approach of Greenberg, for example, a "reasonable fashion" is implicitly defined as never failing two audits in a row. The model developed here extends Greenberg's three-state approach, and involves a different definition of the monitoring probabilities and the introduction of errors of inference.

Consider first, however, a simple version of the repeated play model without error. In what follows it is possible to relax the assumption that all sources are identical without adding to the expositional difficulty. When this is done, it will be convenient to define ρ as a monitoring probability lower than the smallest value of m_i, defined as above for each of the i sources ($m_i = \frac{1}{(1+f_i)(1-\beta) - \alpha f_i}$)

Assume for a moment that Greenberg's results hold, so that every source in group 2 has the incentive to comply. But none in group 1 do. Then group 3 remains empty if no source is assigned to it in the beginning, and the transition probability to group 3 from group 2 is zero. Define the monitoring probability in group 2 to be ρ for every source. Let the monitoring probability group 1 be expressed as a function of ϵ and ρ: $z(\epsilon, \rho)$. In these circumstances the repeated play game produces movement of sources among groups 1 and 2 that is captured by a simple markov process with the following transition matrix (where G_i stands for group i):[6]

$$\begin{array}{ccc} & G_1 & G_2 \\ G_1 & 1-z(\epsilon,\rho) & z(\epsilon,\rho) \\ G_2 & \rho & 1-\rho \end{array}$$

It is easy to solve for the stationary probabilities Π_1, and Π_2, which can be intrepreted as the fractions of sources in the two groups in the long run:

$$\Pi_1 = \frac{\rho}{\rho + z(\epsilon,\rho)}$$

$$\Pi_2 = \frac{z(\epsilon,\rho)}{\rho + z(\epsilon,\rho)}$$

The reason for expressing z as a function of ϵ and ρ is that Π_1, must equal ϵ, where ϵ is the required upper limit on the noncompliance fraction. Thus:

$$\epsilon = \frac{\rho}{\rho + z(\epsilon,\rho)}$$

$$\text{or } z(\epsilon,\rho) = \rho\left(\frac{1}{\epsilon} - 1\right)$$

To guarantee $z < 1$, we must have $\rho < \frac{\epsilon}{1-\epsilon}$, but to guarantee $z < m$ so that in fact the incentive is not to comply in group 1, it is necessary that $\rho < m\left(\frac{\epsilon}{1-\epsilon}\right)$, an obviously stricter requirement. If $\epsilon = \Pi_1$, $1-\epsilon = \Pi_2$ the fraction of sources being audited in the steady state condition is:

$$\varepsilon\left(\rho\left(\frac{1}{\varepsilon}-1\right)\right)+(1-\varepsilon)\rho=2\rho(1-\varepsilon)$$

This is also the budget fraction as defined above for the single play game.

That is: $\frac{N \cdot 2\rho(1-\varepsilon) \cdot M}{N \cdot M} = 2\rho(1-\varepsilon) = r$.

What, then, is the relation between the budget fractions required in single play and multiple play games in the absence of error? Above, the errorless single-play budget fraction, r_{sp}, was found to be $(1-\varepsilon)m$ (footnote 5). Calling the multiple-play budget fraction r_{mp}, it is seen that:

$$\frac{r_{sp}}{r_{mp}} = \frac{(1-\varepsilon)m}{(1-\varepsilon)2\rho} = \frac{m}{2\rho}$$

Therefore, the budget required to produce a compliance fraction at least $(1-\varepsilon)$ will be smaller for the multiple play game as long as $m > 2\rho$ or $\rho < \frac{m}{2}$ (Thus, $p < \min(\frac{m}{2}, \frac{m\varepsilon}{1-\varepsilon})$ is a rule guaranteeing both sensible transition probabilities and budget savings from going to the multiple play game. If $\varepsilon = 1/3$, the two arguments for min ( ) are equal, and for $\varepsilon < 1/3$. the requirement that $\rho < m\varepsilon/(1-\varepsilon)$ governs. Therefore, in situations with small allowed fraction of sources in violation, the ratio $\frac{r_{sp}}{r_{mp}}$ will be greater than or equal to $1/2(\frac{1-\varepsilon}{\varepsilon})$, independently of the other parameters, such as fine size.

Introducing Errors Of Inference Into The Repeated Game

If the monitoring instruments available to the agency are not perfect but display random error (or if the parties cannot control their own actions perfectly but may randomly violate regulations in spite of intending not to) there will again be probabilities both of missing actual attempts to evade the regulations and of incorrectly identifying violations by sources in or attempting to be in compliance. The major implication of this change in assumptions will be that group 3 can no longer be assumed to be empty even if the incentive to comply exists in group 2. Indeed, group 3 will be an absorbing state if it is treated as a perpetual punishment for those caught in violation in group 2. Eventually, becuase of false positives, all parties would be in group 3, and the budget requirement would be 1 if all sources in group 3 are audited every period.

The major alteration to the repeated game structure required to accomodate these errors of inference is the introduction of some possibility of escape from group 3. But to link escape to successful passing of a monitoring visit, while perhaps initially tempting, is dangerous. Consider, for example, a design in which a party in group 3 is monitored each period. If success in a monitoring visit led to release back to group 2, the release probability would be $1-\alpha$, where α is the probability of falsely identifying a violation when compliance is occuring. The average time spent in group 3 by a party would be $1/(1-\alpha)$. It will be shown below that for reasonable values of the other parameters, the escape probability must be much less than $1-\alpha$.

Rather, the probability of escape should be tailored to induce compliance in group 2, given the error structure and the size of the fine levied for discovered violations. Therefore, we write the transition matrix for a repeated game, with three groups and audit errors as:

$$\begin{array}{cccc} & G_1 & G_2 & G_3 \\ G_1 & 1-z(.)(1-\beta) & z(.)(1-\beta) & 0 \\ G_2 & \rho(1-\alpha) & 1-\rho & \rho\alpha \\ G_3 & 0 & E & 1-E \end{array}$$

where, as before, ρ is fixed arbitrarily within an upper limit, and z(.) will be chosen to maintain $\Pi_1 = \epsilon$, the fraction of sources in violation, and will be a function of E, ρ, α, β, as well as ϵ. In what follows, ρ is taken to equal m/k, with k an arbitrary constant. α is the probability of a false positive and β the probability of a false negative, as above.

In this setting it is tedious but straightforward to derive the following results:

(i) $z(.) = \frac{E\rho(1-\alpha)(1-\varepsilon)}{\varepsilon(1-\beta)(E+\rho\alpha)}$ in order that $\Pi_1 = \varepsilon$

(ii) Then, $\Pi_2 = \frac{E(1-\varepsilon)}{E+\rho\alpha}$ and

$$\Pi_3 = \frac{\alpha\rho(1-\varepsilon)}{E+\rho\alpha}$$

(iii) $r = \varepsilon z + \frac{E(1-\varepsilon)\rho}{E+\rho\alpha} + \frac{\alpha\rho(1-\varepsilon)A_3}{E+\rho\alpha}$

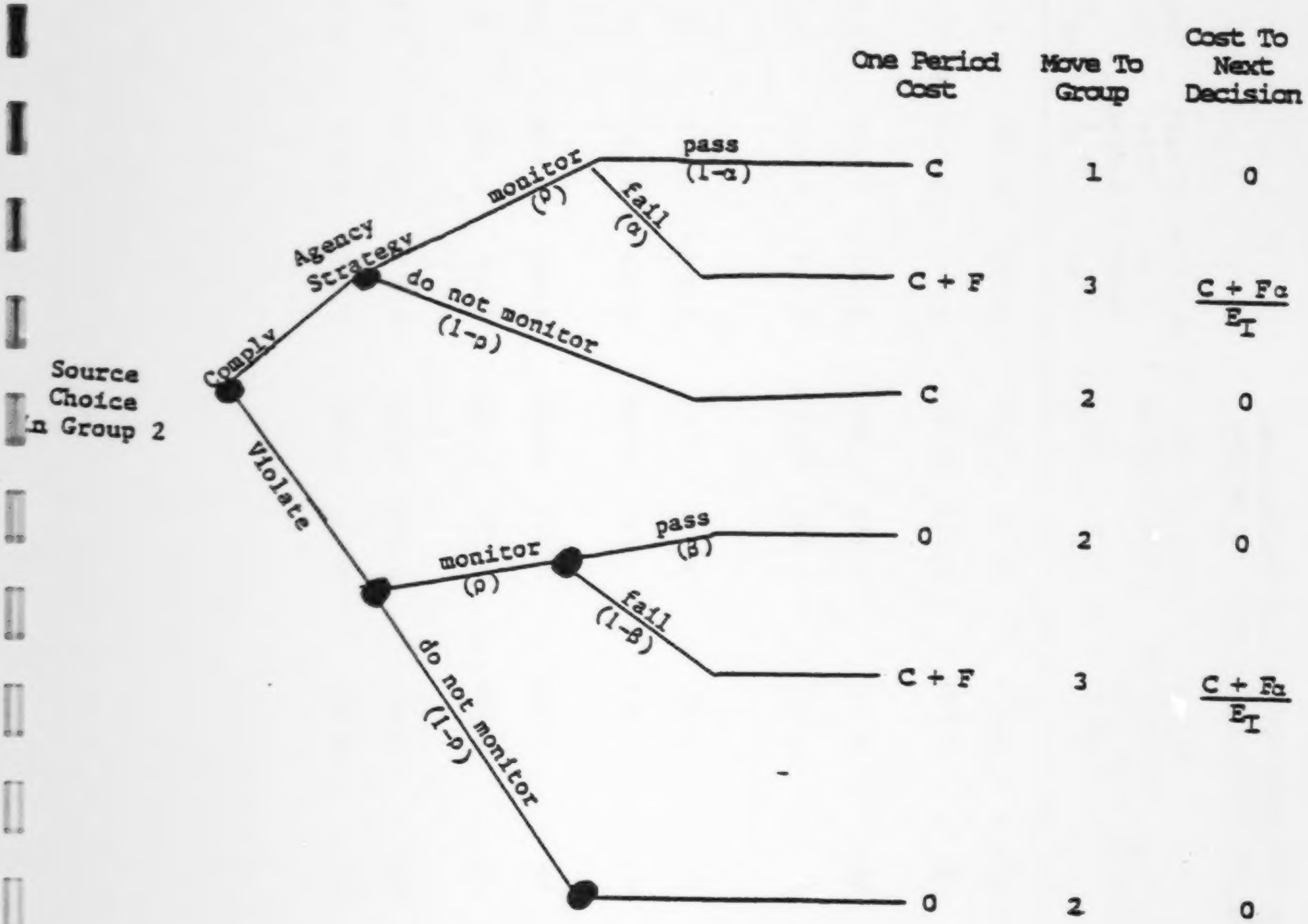
Where A_3 is the monitoring probability in group 3 and can be as low as is consistent with giving each source in that group the incentive to comply in every period. The results reported below generally involve assuming that $A_3 = 1$

(iv) Since $z = \frac{E\rho(1-\alpha)(1-\varepsilon)}{\varepsilon(1-\beta)(E+\rho\alpha)}$ and $\varepsilon z = \frac{E\rho(1-\alpha)(1-\varepsilon)}{(1-\beta)(E+\rho\alpha)}$

$\frac{\partial r}{\partial \varepsilon} < 0$, for r can be written as $(1-\varepsilon)[K]$, where $K > 0$ for

meaningful values of α, β, E, ρ, and A_3, and $\frac{\partial K}{\partial \varepsilon} = 0$

It is possible to derive an upper limit for E, call it E_I, that assures a source will have the incentive to comply in group 2. This derivation may be motivated by displaying the decision tree facing a source in group 2:


One Period Cost
Move To Group
Cost To Next Decision
Source Choice in Group 2
Comply
Violate
Agency Strategy
monitor (ρ)
do not monitor (1-ρ)
pass (1-α)
fail (α)
pass (β)
fail (1-β)
C
1
0
C + F
3
$\frac{C + F\alpha}{E_I}$
C
2
0
0
2
0
C + F
3
$\frac{C + F\alpha}{E_I}$
0
2
0

In this situation, the expected cost of compliance is:

$$E(\text{compliance}) = \rho\ [(1-\alpha\ )C + \alpha(C+F)+\alpha\ (\frac{C+F\alpha}{E_I})]+(1-\rho\ )C;$$

While the expected cost of violation is: $E(\text{violation}) = \rho(1-\beta\ )[(C+F)+\frac{C+F\alpha}{E_I}]$;

Requiring that E(compliance) < E (violation) implies that:

$$E_I < \frac{\rho(1+f\alpha)(1-\beta-\alpha)}{1+\rho(f\alpha-(1-\beta)(1+f))}$$

Consistent values of ρ, f, α, β, and ε are used in table 3 to derive values of E_I. It may seem counter intuitive that as fines grow relative to avoidable costs of compliance, the upper limit on E_I falls (the time spent in group 3 grows). This occurs because as fines are increased, m and hence, ρ, the monitoring probability for group 2 both fall. This reduces the chance of

TABLE 3: CHARACTERISTICS OF THE REPEATED-PLAY MONITORING GAME, INCLUDING RELATIVE BUDGET SIZE, AS FUNCTIONS OF BASIC PARAMETERS

Required fraction in non-compliance ($\epsilon = \Pi_1$)	Relative fine (f)	Type I error (α)	Type II error (β)	K ($\rho = \frac{m}{k}$)	Escape probability from group 3 (E_I)	Size of group 2 (Π_2)	Size of group 3 (Π_3)	Multiple play budget (r_{mp})	Single play budget (r_{sp})	Budget ratio (r_{sp}/r_{mp})
0.2	1.2	0.05	0.2	40	0.012	0.754	0.044	0.071	0.493	6.96
0.2	3.2	0.05	0.2	40	0.007	0.757	0.043	0.062	0.263	4.21
0.1	1.2	0.05	0.2	40	0.019	0.849	0.050	0.080	0.554	6.91
0.1	3.2	0.05	0.2	40	0.007	0.851	0.049	0.071	0.296	4.18
0.2	1.2	0.01	0.01	40	0.012	0.797	0.008	0.024	0.372	15.8
0.2	3.2	0.01	0.01	40	0.006	0.794	0.008	0.018	0.196	11.1
0.1	1.2	0.01	0.01	40	0.012	0.890	0.008	0.026	0.418	15.8
0.1	3.2	0.01	0.01	40	0.006	0.892	0.008	0.019	0.220	11.6
0.2	1.2	0.05	0.2	200	0.002	0.753	0.047	0.052	0.493	9.48
0.2	3.2	0.05	0.2	200	0.001	0.759	0.043	0.046	0.263	5.72

getting into group 3 at all, an effect that is balanced by the extension of expected time there. This demonstrates the central place in this system of the threat of residence in group 3 and the subsidiary role of the size of the fine levied for violation.

It is possible to substitute E_I into the expression for the budget, r, and thus to obtain an expression for the budget in terms of the underlying parameters, consistent with the requirements (i) that the proportion of violators be kept to ϵ; and (ii) that parties in group 2 have an incentive to comply. Unfortunately, the resulting expression is rather messy and does not bear any simple relation to the budget expression for the single play version of the game with error. Thus we obtain for r_{mp} (with error):

$$r_{mp}\ \text{(with error)} = \frac{(1-\epsilon)\rho\left[\left(\frac{\rho(1+f\alpha)(1-\beta-\alpha)}{1-\rho/m}\right)(2-\alpha-\beta)+\alpha(1-\beta)\right]}{\left(\frac{\rho(1+f\alpha)(1-\beta-\alpha)}{1-\rho/m}+\rho\alpha\right)(1-\beta)}$$

$$= \frac{(1-\epsilon)\rho\left[E_I(2-\alpha-\beta)+\alpha(1-\beta)\right]}{(E_I+\rho\alpha)(1-\beta)}$$

where monitoring in group 3 is taken to be certain; $A_3 = 1$ in earlier notation.

To provide some feel for the budgetry advantage of using the threat implicit in the multiple-play and grouping approach to the enforcement game, however, table 3 also shows budget values for two different allowed proportions of violators, two levels of fines in relation to avoidable compliance costs, and two sets of error probabilities. The ratio of single-play to multiple-play budgets, parameter choices equal, is a measure of the improvement bought by introducing record-dependent monitoring probabilities. Note that with small errors, the ratio is over 15 to 1. Or, said another way, for this range of parameter choices, the budget required for a multiple-play scheme would be between 6 and 25 percent of a single-play approach designed to achieve the same compliance results.

Several other observations about table 3 are in order:

- As expected, cet. par. the required budget is lower when the allowed fraction of parties in violation is higher.
- But notice that the ratio of the single play to multiple-play budgets appears to be independent of the allowed fraction in violation. In fact, the small differences found in table 3 reflect rounding errors, for it is straightforward to show that in the ratio, all terms involving ε cancel.
- When the error structure is improved (both α and β reduced) budget requirements are reduced, other specifications remaining the same.
- A given increase in the size of the fine relative to avoidable cost has a larger effect on required budget in the single-play model than it does in the repeated game. This difference is more pronounced the less favorable the error structure.
- Budget size is affected by choice of the arbitrary constant, K, used to translate m (the compliance-inducing monitoring probability) into (the monitoring probability actually applied to group 2). It appears that simply by choosing K larger and larger (monitoring less and less in group 2) the budget can be driven closer and closer to zero. But this is an illusion created by the small range covered in the table. In fact, as K goes to infinity, the stationary probabilities Π_2 (fraction of sources in group 2) and Π_3 (fraction of sources in group 3) can be approximated by:

$$\Pi_2 \approx \frac{1-\varepsilon}{1-\alpha}$$

$$\Pi_3 \approx \frac{\alpha(1-\varepsilon)}{1+\alpha}$$

Since both $\rho (= \frac{m}{k})$ and $z (= \frac{1-\varepsilon}{\varepsilon} (\frac{E\rho (1-\alpha)}{(1-\beta)(E+\rho\alpha)})$ go to zero as K goes to infinity, in the limit the budget depends only on Π_3. When the monitoring probability in group 3 is taken to be 1, as it is here, the repeated play budget approaches $\frac{\alpha(1-\varepsilon)}{1+\alpha}$ in the limit. Therefore, the budget cannot be driven to zero by choosing an arbitrarily large K. And the limiting value of the ratio r_{sp}/r_{mp} is $(\frac{m}{1-\alpha})(\frac{1+\alpha}{\alpha})$. For example, for the first case in table 3, where $\varepsilon = 0.2$ and $\alpha = 0.05$,

$$\lim_{k \to \infty} \Pi_3 = \lim_{k \to \infty} r_{mp} = \frac{0.5(0.8)}{1.05} = 0.038$$

and $$\lim_{k \to \infty} (r_{sp}/r_{mp}) = \frac{0.585}{0.95} \cdot \frac{1.05}{0.05} = 12.93.$$

- Finally, higher fines imply a smaller budget advantage for multiple-play over single-play structures and the larger the errors, the smaller the advantage for constant fine size.

Concluding Comments

This paper makes a preliminary case for using game theory notions, especially the repeated game model, in the design of monitoring and enforcement systems. Most importantly, desired levels of compliance may be obtained with smaller budget allocations for monitoring when information on past compliance is used to define future monitoring probabilities. The relevance of these results for actual policy will depend on the resolution of several questions, some involving law and some extensions of the economics considered here.

The most important legal point is likely to be whether or not an arbitrary probability of escape from group 3, E, can in fact be applied to pollution sources (or any other monitored parties).

By far the most important economic question is how the results derived above are affected by the introduction of discounting at a positive interest rate. There is no question that discounting will reduce the effectiveness of the scheme, for the threat of a given stay in group 3 will be lower, everything else equal. The question is, by how much? Will the required monitoring budget, given a required upper limit on noncompliance, an error structure, and a fine size, be less than that implied in a single-play approach? Will it be enough less to make the relatively elaborate grouping mechanism worthwhile?

FOOTNOTES

1. It is an interesting and challenging feature of air and water pollution discharges that they are fugitive events in the sense that they must be measured as they occur or they cannot be measured at all. (There are statistical techniques for inferring air pollution discharges from ambient pollutant concentrations. (See Gordon 1980). That is, enforcing pollution control permit terms is more closely analogous to enforcing auto speed limits than, say, the tax laws. The situation for containerized hazardous wastes is different, of course.

2. In fact, most monitoring visits are the result of the application of rules of thumb such as: Visit large sources once every N months. Such visits usually also involve prior announcement for convenience and to avoid legal battles over rights of entry. Thus, the deterence effect of the visits is even less than that implied by the rates.

3. One visit is assumed to be the maximum required to guarantee compliance in a period (subject to the qualifications implied by errors of inference). That is, if a monitoring visit occurs and a violation is detected, that violation is assumed to be corrected for the remainder of the period. It is further assumed that the full costs of compliance in the period, C, will be incurred if a source in violation is caught and forced into compliance. A false violation notice (false positive) implies that a fine is levied on a source already spending C to comply.

4. A natural question is, why not announce that m applies to all sources, but actually only apply r? As a short-term strategy of deceit this may seem appealing. But it is conceptually weak reed on which to build a policy, for assuming that deceit can successfully be practiced indefinitely seems to imply zero information gathering activity and inferential ability on the part of the sources. Similarly, relying on an announced monitoring probability, m, applying to an unknown fraction of sources $\frac{N_r}{N}$ gives an effective monitoring probability for a source of:

$$r = \frac{N_r . m}{N} = \frac{\frac{N.r.m.}{m}}{N} .$$

It seems wise to assume that each source will infer the actual probability it faces, and therefore no sources will have the incentive to comply.

5. If $\alpha = 0$, the measurements are effectively perfect for these purposes and $r = (1 - \varepsilon)m$.

6. Proving that it is optimal for a source to comply in group 2 while violating in group 1 involves defining the following: $E(V_2)$ is the expected value of the strategy violating when in group 2; $E(C_2)$ is the expected value of the strategy of complying in group 2, and $E(V_1, C_2)$ in the expected value of violating in group 1 and complying in group 2. Observe the following:

$$E(V_1, C_2) = (1 - z)0 + zE(C_2)$$

$$E(C_2) = C + \rho(E(V_1, C_2)) + (1-\rho)(E(C_2))$$

$$E(V_2) = 0 + \rho \sum_{t=1}^{\infty} C_t + (1 - \rho)E(V_2)$$

so that:

$$E(V_2) = \sum_{t=1}^{\infty} C_t \to \infty$$

$$E(C_2) = C + \rho(zE(C_2)) + (1 - \rho)E(C_2)$$

$$E(C_2)[1 - \rho z - (1 - \rho)] = C$$

$$E(C_2) = \frac{C}{\rho - \rho z}$$

And so long as $z < 1$, this number is positive but finite, so that $E(C_2) < E(V_2)$. $z < 1$ is guaranteed by the rule proposed in the text for defining z in terms of ε and ρ.

REFERENCES

Becker, Gary S. 1968. "Crime and Punishment: An Economic Analysis," *Journal of Political Economy*, 76 (March/April), pp. 169-217.

Casey, Don, Peter M. Nemetz, and Dean H. Uyemo. 1983. "Sampling Frequency for Water Quality Monitoring: Measures of Effectiveness", *Water Resources Research*. vol. 19. no. 5. (October) pp. 1107-1110.

Fortuna, Ricard C. and David J. Lennett. 1987. *Hazardous Waste Regulation: The New Era.*, (McGraw Hill, NY).

Gordon, Glen E. 1980. "Receptor Models" *Environmental Science and Technology*. vol. 14, no. 1, pp. 792-800.

Greenberg, Joseph. 1984. "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach," *Journal of Economic Theory*, vol. 32, no. 1, (Feb.), pp. 1-13.

Landsberger, Michael, and Isaac Meilijson. 1982. "Incentive Generating State Dependent Penalty System," *Journal of Public Economics*, vol. 19, pp. 333-352.

McKean, Roland. 1980. "Enforcement Costs in Environmental and Safety Regulation," *Policy Analysis*, vol. 6, no. 3, pp. 269-289.

Russell, Clifford S. 1982. *Pollution Monitoring Survey Summary Report*, (Washington: Resources for the Future).

Russell, Clifford S. 1987. forthcoming, "Monitoring and Enforcement," in Paul R. Portney, (ed.) *Environmental Regulation In The U.S.: Public Policies and Their Consequences*. (Washington, DC: Resources for the Future).

Russell, Clifford S., Winston Harrington and William J. Vaughan. 1986. *Monitoring and Enforcing Pollution Control Laws*. (Washington, DC: Resources for the Future).

U.S. GAO, 1979. *Improvements Needed in Controlling Major Air Pollution.* (Washington, DC: U.S. GAO) January, CED-78-165.

U.S. GAO, 1980. *Costly Wastewater Treatment Plants Fail to Perform as Expected.* (Washington, DC: U.S. GAO) November CED-81-9.

U.S. GAO, 1983. *Wastewater Dischargers Are Not Complying With EPA Pollution Control Permits*. (Washington, DC: U.S. GAO) December GAO/RCED-84-53.

U.S. GAO, 1985. *Air Pollution: Environmental Protection Agency's Inspections of Stationary Sources*. (Washington, DC: U.S. GAO) Oct. GAO/RCED-86-1BR.

U.S. Nuclear Regulatory Commission, Office of Inspection and Enforcement. 1983. Inspection and Enforcement Manual, chapter 0300 (January) Washington, DC.

Wasserman, Cheryl. 1984. "Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies: United States: A National Review," prepared for the Organization for Economic Co-operation and Development, Environmental Directorate, (October). Washington, DC.

The Behavior of the Firm When Facing Uncertain Enforcement of *Ex Ante* Environmental Regulation and Tort Liability Standards

by

Gary V. Johnson
Institute for Environmental Studies
and
Department of Agricultural Economics

and

Thomas S. Ulen
Department of Economics
and
Institute of Government and Public Affairs
University of Illinois at Urbana-Champaign

Prepared for the AERE Workshop on
Environmental Enforcement and Monitoring
Newark, Delaware
August, 1987

Acknowledgements: We wish to thank Juha Alho, Institute for Environmental Studies, University of Illinois at Urbana-Champaign for his comments on an earlier draft and the Research Board of the University of Illinois for their generous financial assistance.

The economic policy analysis of environmental issues has generally focused on the relative efficiency in minimizing environmental harms of the unregulated market and of administrative agency (or what we call *ex ante*) regulation. Once the scholarly analysis has indicated the circumstances that dictate governmental intervention in the market, it has then focused on the relative efficiency of different forms of this *ex ante* regulation. For example, the standard reliance on a system of fixed fines for violation of environmental regulations has been challenged as inefficient by comparison to a system of sliding effluent fees or to a requirement that potential polluters install the best available (in the sense of least-polluting) technology for production of their output.

Recently, economists and lawyers familiar with economics have explored the efficiency of exposure to tort liability (what we will call *ex post* regulation) as part of a policy for minimizing external costs such as environmental harms.[1] The premise is that potential polluters can be constrained to produce an efficient amount of pollution by making them liable in a private cause of action for compensating those whom they have harmed. Until very recently this method of regulating environmental harms has been a rare exception.[2]

This article attempts to bring together these two strands of the environmental and the law and economics literatures in order to investigate the relative efficiency of *ex ante* and *ex post* regulation. The question naturally raised by this conjunction of the regulatory literature is whether *ex ante* and *ex post* regulation are complements or substitutes. Educated economic intuition would suggest that they are substitutes; either alone would seem to be able to accomplish the goal of creating efficient incentives but to use them jointly would seem to be wasteful duplication. It turns out that this intuition is correct only when there is no uncertainty in the enforcement of either the administrative agency standard or the tort liability standard. In that unlikely case, the two methods of regulation are, indeed, substitutes, and the social costs of regulating the externality are minimized by relying exclusively on whichever of the two forms of regulation has the lower administration and enforcement costs. The far more likely case is that there is uncertainty in the enforcement of either or both forms of regulation (Russel *et al.*). Where that is true, then *ex ante* and *ex post* regulation become complementary regulatory tools. We have recently shown that the joint use of these forms of regulation creates efficient incentives for potential externality-generators when there is certainty regarding the enforcement of the *ex ante* regulatory standard but so much uncertainty regarding the enforcement of the *ex post* liability rule that firms take less precaution to prevent environmental harms than is socially optimal (Johnson *et al.*). In this article we extend those results by considering the case where there is uncertainty in the enforcement of *both* forms of regulation.

---

[1] See Robert D. Cooter and Thomas S. Ulen, *Law and Economics* (Scott, Foresman and Co.: Glenview, IL, 1987), pp. 326-371.

[2] For exceptions, see Boomer v. Atlantic Cement Co., Inc., 26 N.Y.2d 219, 309 N.Y.S.2d 312, 257 N.E.2d 870 (Court of Appeals of New York, 1970), and Spur Industries, Inc. v. Del E. Webb Development Co., 494 P.2d 701 (Ariz. 1972). The economics of the cases are discussed in Cooter and Ulen, op. cit., 171-185.

The policy conclusions that result from analyzing the effect of uncertain enforcement in both forms of regulation can be readily stated. The existence of uncertain enforcement of both an *ex ante* administrative agency regulation and an *ex post* liability rule for a given environmental externality strengthens the case for the simultaneous use of these two forms of regulation but slightly alters the relationship between the two. When only the enforcement of the *ex post* standard is uncertain, the (certain) *ex ante* standard should, surprisingly, be set below the socially optimal level of care. When there is uncertainty in the enforcement of both standards, it is still true that the *ex ante* standard should be set below the *ex post* standard, which, by assumption, is set equal to the socially optimal level of care. But now the distance between the two standards depends on the relative degree of uncertainty in the enforcement of the two standards in an intuitively plausible way that we explain below.

The remainder of the paper is organized in the following way. In Section II we discuss the economic analysis of tort liability and apply that analysis to the policy issues created by the generation and disposal of hazardous wastes. Section III develops the formal model of the firm's behavior when faced with both *ex ante* and *ex post* regulation of potential environmental harm with special attention to the impact of the uncertain enforcement of both of those standards on the potential polluter's behavior. The paper concludes with remarks on two matters. First, we suggest the implications of the model for public policymaking regarding environmental harms. Second, we speculate on the next steps in the theory of integrating *ex ante* and *ex post* regulation and on the empirical work that our formal model suggests.

II. The Economics of Tort Liability

The economic analysis of tort liability standards is sufficiently new and arcane that a brief introduction is in order. Recall that the simple economic premise is that the potential injurer and victim will be induced to take optimal precaution against harm if they are liable for compensating those whom they harm. This potential liability becomes a part of the decisionmaker's anticipated costs, which he then attempts to minimize by taking the optimal amount of precaution. The common law recognizes two different tort liability standards—negligence and strict liability. But the law has not provided a convincing explanation of the circumstances in which negligence is superior to strict liability and *vice versa*. The economic analysis of tort liability *has* provided such an explanation. Let us briefly summarize this explanation. It is an important one in designing public policy for environmental harms and in introducing the formal model of the next section.

The tort liability standard of negligence establishes a legal standard of care or precaution that the potential injurer and victim owe to one another.[3] If a party violates the legal standard by taking less care or precaution than he owes to the other party and that failure to take care proximately caused the harm, then he is negligent or at fault and is liable for the losses suffered by the victim; if he takes at least as much care as required by the legal standard or if his actions did not proximately cause harm, even though he was negligent, then he is not negligent or at fault and is not liable for the victim's losses. The negligence standard comes in several varieties—simple negligence, negligence with contributory negligence, and comparative negligence—that differ in the extent to which they take into account the victim's own precaution against the occurrence of a harm. The general conclusion of the literature on the efficiency aspects of the negligence rule is that under certain conditions all forms of the negligence rule create equally efficient incentives for potential victims and injurers to take precaution and, therefore, minimize the social costs of external harms. The conditions under which this conclusion holds are that (1) the legal standard of care is set equal to the socially optimal level of care, (2) precaution is "bilateral," and (3) there is no imperfection or uncertainty in determining whether or not a party complied with the legal standard of care. By "bilateral precaution" we mean that the technology of precaution is such that both parties may reasonably take action to reduce the probability or severity of the external harm. An example would be automobile accidents: typically there is something that both drivers can do to reduce the probability or severity of an accident, even if in any given accident only one of them was in fact at fault. A point worth emphasizing here is that *all* forms of the negligence rule are equally efficient when there is no uncertainty in the enforcement of the legal standard of care.

---

[3] The following material draws on Cooter and Ulen, op. cit., 326-371.

When we introduce uncertain enforcement into the economic analysis of negligence, the conclusion changes. Suppose that enforcement is uncertain in the following sense: it is possible that (1) a party who complied with the legal duty of care is nonetheless held liable or (2) a party who failed to comply with the legal duty of care is not held liable. Under this uncertainty there may be important efficiency differences among the forms of the negligence rule (Cooter and Ulen, 1986; Craswell and Calfee, 1984 and 1986.). This highlights the importance of uncertainty in modeling behavior under *ex post* regulation.

The other great liability standard is strict liability. Under that standard, an injurer is liable for the victim's losses if he proximately caused the harm. The potential injurer is under an absolute duty not to cause harm; there is no legal standard of care or precaution that he may take to exonerate him from liability for the victim's losses. Moreover, there is usually no requirement that the potential victim himself take care; the entire burden is on the potential injurer. Strict liability creates efficient incentives for precaution by the potential injurer under three conditions: (1) when he is liable to the victim for "perfect" compensation, (2) when precaution is "unilateral," and (3) when there is no uncertainty in the enforcement of the strict liability standard. Perfect compensation is an award of money damages to the victim that makes him indifferent between the state of having suffered the harm but receiving the money damages and that of never having suffered the harm. Precaution is said to be unilateral when only the potential injurer may reasonably take action to reduce the probability or severity of harm. Uncertain enforcement can arise under strict liability in two ways. First, the determination of proximate cause may be subject to uncertainty: an injurer may be held strictly liable when in fact his actions did *not* proximately cause harm or he may be excused from liability when in fact his actions *did* proximately cause the injury. Second, the victim's compensation can be imperfectly measured. Potential injurers thus may be liable for widely-varying amounts of compensation: some victims whose loss was minimal may be vastly over-compensated while others whose loss was extraordinarily large may be under-compensated. Where there is uncertain enforcement in either or both of these senses, the efficiency of strict liability is lessened.

The uncertain enforceability of tort liability standards clearly lessens their efficiency. (We assume that the other conditions specifying whether strict liability or negligence is the appropriate liability rule are met; that is, the only source of inefficiency upon which we wish to focus is uncertain enforcement; for instance, strict liability is not being applied in circumstances of bilateral precaution, nor is negligence being applied in cases of unilateral precaution.) The question that this lessened efficiency raises is whether there is some way to improve the efficiency of an uncertainly-enforced *ex post* regulatory rule. There are several possibilities:

> (1) depending on the direction of the inefficiency (i.e., whether it leads to too much or too little precaution), the rules for establishing fault or causation could be relaxed or tightened;
>
> (2) assuming that there is so much uncertainty that potential injurers inevitably take too little precaution, courts could routinely award punitive damages in addition to compensatory damages) in those instances where an injurer is held liable;
>
> (3) if the inefficiency is extraordinarily large, some alternative regulatory tool for minimizing the external harm might be substituted (e.g., the activity could simply be outlawed or victims could be compensated in an administered compensation or no-fault system);
>
> (4) a complementary regulatory tool could be used (e.g., in addition to exposing potential wrongdoers to tort liability, they could also be exposed to siting requirements or quality controls backed by fines.

To focus the discussion let us apply the foregoing analysis to a concrete problem in environmental regulation: harms arising from exposure to hazardous wastes. Let us first briefly summarize, for future use, the federal and state statutory regulations dealing with that problem and then discuss how the economic analysis of tort liability might be applied to this issue.

Federal statutory regulation of hazardous wastes began with the Safe Drinking Water Act of 1974.[4] Among other things that Act told the US. Environmental Protection Agency (EPA) to require the states to promulgate regulations dealing with the underground injection of hazardous wastes. Two years later Congress passed the Resource Conservation and Recovery Act of 1976 (RCRA).[5] That Act provided for tracking hazardous wastes from the time and point of generation until final disposal but failed to provide for any problems associated with improper disposal before 1976. In 1980 Congress attempted to correct this failure in the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA).[6] That Act empowered the Federal government through the EPA to impose liability for cleaning up hazardous waste disposal sites closed before the enactment of RCRA.[7] CERCLA has been revised but its basic structure remains intact.

State statutory regulation is predictably less coherent. Some states have "mini-superfunds,"[8] and many have additional, more direct controls. For example, Illinois and Massachusetts have banned landfilling, the most popular and economical method of disposing of hazardous wastes;

---

[4] 42 U.S.C. secs. 300f-300j (1982).

[5] 42 U.S.C. secs. 6901-6907, 6911-6916, 6921-6931, 6941-6954, 6961-6964, 6971-6979, 6981-6986 (1982).

[6] 42 U.S.C. secs 9601-9657 (1982). CERCLA created the "Superfund," a fund financed by Congressional appropriations and taxes on those who produce hazardous and toxic wastes. The EPA is empowered to use this fund to clean up dangerous waste sites and then to bring actions against various parties for recovery of the cleanup costs under sec. 107 of the Act.

[7] For an excellent summary of the law on hazardous wastes, see "Developments in the Law -- Toxic Waste Litigation," 99 *Harv. L. Rev.* 1459 (1986).

[8] For an example see the Minnesota Environmental Response and Liability Act (MERLA), Minn. Stat. Ann. §§115B.01-115B.24 (West 1985 supp.).

some states have extended the reporting requirements of RCRA to small firms (such as dry cleaners and gasoline stations) that are exempt from the federal regulations.

The common law tort liability treatment of harms arising from the generation and disposal of hazardous wastes is in its infancy. Only a handful of actions have been filed so that it is not clear how the common law will treat these matters. For that reason we may begin at the beginning and explore how the common law *should* deal with hazardous wastes by applying the economic analysis of tort liability sketched above.

Consider how the distinction between negligence and strict liability applies to the environmental harms that might arise from the generation or disposal of hazardous wastes. For those harms, precaution is certainly unilateral in the sense that disposers and generators are the only parties to whom society may reasonably look for actions that will reduce the probability or severity of harm from those sources. Thus, if this harm is to be regulated using an *ex post* liability rule, injurers (the disposers and perhaps the generators of hazardous wastes) should be held strictly liable to those they have harmed.

But what about the problems of uncertain enforcement under strict liability?[9] Recall that in making a claim for recovery under strict liability, the victim must show that the injurer proximately caused the harm (but not that in doing so the injurer violated a standard of due care owed to the plaintiff). Two problems are likely to arise in establishing that the generation or disposal of hazardous waste has proximately caused a harm. First, the scientific evidence on causation between exposure to these substances and personal or property injury is at an early stage of development. Thus, it may well be the case that although by the standards of the academic community the causal connection may be reasonably well established, by the standards of proof required in a court the causal connection is not clear enough to permit recovery in amounts close to perfect compensation (Farber, Davis). Second, harms arising from the generation or disposal of hazardous wastes may not become manifest for long periods of time, sometimes more than a generation. If so, the evidence necessary to establish proximate cause for recovery under a strict liability theory may be so distant or so clouded that otherwise meritorious plaintiffs cannot recover. Where this is the case, then the generators and disposers of hazardous wastes do not receive the appropriate signal from the tort liability system about the appropriate level of precaution to take, and as a consequence they may take too little precaution.

---

[9] The following discussion comes from Gary V. Johnson and Thomas S. Ulen, "Designing Public Policy Toward Hazardous Wastes: The Role of Administrative Regulations and Legal Liability Rules," *American Journal of Agricultural Economics* (Dec. 1986) pp. 1269-1270.

To what extent can the strict liability standard be amended to take account of these special problems of recovering for harms arising from the generation and disposal of hazardous wastes? With regard to the problem of establishing proximate cause where scientific evidence is weak, there are several possibilities. Legislation might relax the plaintiff's burden of proof on the causation issue. This is the tack adopted in the 1983 Minnesota Environmental Response and Liability Act (MERLA) and proposed by some commentators for widespread adoption.

This relaxation of the plaintiff's burden of proof on the causation issue for harms arising from the generation or disposal of hazardous and toxic wastes may cure the first problem noted above, but it does so at a high cost. Tampering with the traditional causation standard is a radical step that requires extraordinary justification, a justification that no one has offered. Naturally the question arises: If the inadequacy of scientific knowledge to establish proximate cause in the case of harms arising from hazardous and toxic wastes justifies relaxing the plaintiff's burden of proof in those harms, why may we not also relax plaintiffs' burdens in all other instances of inadequate scientific knowledge? Thus far, there has been no good answer to that question. And that fact suggests that relaxation of the plaintiff's causation requirement is not yet an acceptable method of making strict liability for harms inflicted by exposure to hazardous and toxic wastes more efficient.

The second source of uncertain enforcement of strict liability for harms associated with hazardous wastes is the long time lag between exposure and manifestation of the harm. This lag complicates the injurers ability to demonstrate proximate cause and, therefore, makes his recovery, even if his case is meritorious, less likely. This problem has arisen in several well-known modern cases, e.g., in the diethylstilbestrol (DES) and asbestos cases. Some commentators have suggested that the tort liability system can be reformed to accommodate the peculiar evidentiary problems of time-delayed harms by allowing probable victims to recover from a probable injurer before any actual harm has become manifest. The proposal is that where the probability of any harms developing in the future is above some minimum threshold, the potential victim should be allowed to recover the expected damages discounted by the probability that the harm will arise (Cooter and Ulen, 1987).

It is argued that the benefit of allowing recovery under a theory of probabilistic causation for inflicting tortious risk is that the signal to take efficient precaution will be transmitted relatively quickly to potential injurers. Not allowing recovery in these circumstances, it is argued, will greatly reduce the number of cases that plaintiffs can win in the distant future and consequently will greatly dilute the signal to injurers to take efficient precaution.

But the costs of revamping tort law to allow for probabilistic causation and recovery for infliction of tortious risk are also high. One of the most fundamental precepts of tort law is that a harm must have occurred; simply creating a dangerous condition--what has been called, in a famous phrase, "negligence in the air"--is not a sufficient basis ford bring an action. There are good efficiency reasons for limiting recovery in tort to cases of actual harm. Moreover, there are almost insurmountable problems involved in specifying the threshold probability of harm that would trigger liability.

These observations suggest that the uncertain enforcement that is likely to arise under a strict liability standard for harms arising from exposure to hazardous wastes cannot be easily corrected with the *ex post* regulation system itself.

It is still an open question whether an uncertainly-enforced tort liability standard is best supplemented or replaced by uncertainly-enforced federal and state administrative agency regulation and like those we described above. We turn to that question in the next section with the help of a formal model of firm behavior under uncertainly-enforced *ex ante* and *ex post* regulation.

III. Modeling the Firm's Behavior Under Uncertain Enforcement of Regulation

To examine the effect on the firm's behavior of uncertainty regarding the enforcement of both *ex ante* environmental regulation and *ex post* liability, we will first look at a simple model of firm behavior and enrich the model by adding greater complexity. We begin our model development by reviewing a deterministic model of the firm behavior when faced with strict liability that is common to the law and economics literature (Cooter and Ulen, 1987). Strict liability is chosen because of the unilateral nature of precaution regarding many environmental harms. Uncertainty regarding the enforcement of a strict liability rule will then be introduced. We will examine the effects of assumptions concerning the distribution of that uncertainty and the nature of the cost function associated with precaution. Next, we will focus on *ex ante* regulation, beginning with a simple deterministic model of a regulatory standard. Uncertainty regarding the enforcement of the standard will then be introduced and its impacts explored. Finally, we model the firm's behavior under the joint use of *ex ante* regulation and *ex post* liability when enforcement of both forms of regulation is uncertain.

A. A Model of the Firm Facing Strict Liability

Let x be the level of the firm's precaution in preventing an environmental harm. The costs of precaution are given by the function $C(x)$, which is upward sloping, i.e., $C'(x) > 0$, and convex over the relevant region. The expected size of the harm that the firm can anticipate is given by the function $A(x)$, which is assumed to be convex and downward sloping over the relevant region, i.e., $A'(x) < 0$. [$A(x)$ embodies both the severity of the harm and the probability of its occurring.]

The expected social costs of the externality or harm are the sum of the precautionary costs, $C(x)$, and the expected costs of the harm, $A(x)$. Assume that this sum, $[C(x) + A(x)]$, is strictly convex. The socially optimal amount of precaution, x^*, is that which minimizes these social costs, i.e.,

$$\min_x SC(x) = [C(x) + A(x)], \tag{1}$$

where $SC(x)$ is the total social cost. At x^* the marginal expected cost of precaution equals the negative of the expected marginal cost of the harm, i.e.,

$$C'(x^*) = -A'(x^*). \tag{2}$$

It can be shown that in the absence of uncertainty, with perfect victim compensation and unilateral precaution, the firm subject to strict liability will choose the socially optimal amount of precaution, x^* (Cooter and Ulen, 1987).

As we noted informally in the previous section, if uncertainty exists in the determination of proximate cause or in the determination of perfectly compensatory damages for the victim, then the firm may choose a socially inefficient amount of precaution (Johnson *et al.*). To simplify the modeling of enforcement uncertainty regarding the firm's choice of the level of precaution, we assume for the moment that victim compensation is perfect. Given this assumption, the focus becomes the logically prior issue of causality. The firm perceives enforcement uncertainty regarding whether the court will find that its actions proximately caused the victim's harms. This could arise, as noted above, from a lack of the scientific knowledge necessary to establish proximate cause. Let $q(x)$ be the firm's subjective probability distribution around what it perceives as its cost-minimizing level of precaution, $\hat{x}$.[10] Assume that $q(x)$ is a continuous probability density with support $[0,\infty)$. The probability of the firm's actions being deemed the proximate cause of the harm and of the firm's therefore being held strictly liable to the victim is thus given by

$$R(x) = \int_{x}^{\infty} q(x)dx. \tag{3}$$

That is, $R(x)$ is the probability that the firm will be held liable to pay (perfectly compensatory) damages, given a precaution level of amount x.

How will the firm respond to this form of uncertainty? The firm's objective function can be formulated as the minimization of its precautionary and expected liability costs:

$$\text{Total Costs} = TC(x) = C(x) + A(x)R(x). \tag{4}$$

Previously we assumed that $[C(x) + A(x)]$ is strictly convex. Now we assume that $[C(x) + A(x)R(x)]$ is strictly convex. The essence of the model is presented in Figure 1. The socially optimal level of precaution, x^*, is where the marginal precaution costs just equal the negative of the marginal expected harm. When certainty exists in the determination of proximate cause, the firm's optimal amount of precaution will coincide with the social optimum, x^*. But where uncertainty exists in determining causality, then the probability that the firm will be held liable at any given level of precaution, say $\tilde{x}$, is the area $R(\tilde{x})$ under the distribution and to the right of $\tilde{x}$.

---

[10] For the sake of simplicity, we will assume that $\hat{x}$ and x^*, the social-cost-minimizing level of precaution, are equal. Later we will analyze the case where $\hat{x}$ and x^* differ.

It can be shown that if $TC(x)$ is strictly convex, this function has a unique minimum (Johnson *et al.*). The first-order condition for the minimization is

$$C'(\tilde{x}) + A'(\tilde{x})R(\tilde{x}) - A(\tilde{x})q(\tilde{x}) = TC'(\tilde{x}) = 0. \tag{5}$$

Equation 5 is basic to the subsequent analysis and thus deserves some interpretation. The first term on the left-hand-side of the equation is the marginal cost of providing a unit of precaution. The second and third terms sum to the expected marginal liability costs of a unit of precaution and consist of two effects. The first effect, $[A'(\tilde{x})R(\tilde{x})]$, is the marginal expected cost of the harm times the probability of being held liable for the harm if the firm's level of precaution is equal to $\tilde{x}$. This term, which we call the "injury effect," is negative because $A'(\tilde{x})$ is negative (more precaution reduces the severity and probability of harms) and $R(\tilde{x})$ is always positive. But there is also a savings from providing slightly higher precaution, a reduction in the probability of being held liable. This savings is captured in the term $[-A(\tilde{x})q(\tilde{x})]$, which we call the "liability effect." This term is negative because $-A(\tilde{x})$ is negative and, by assumption, $q(\tilde{x})$ is positive. It follows that the firm minimizes its expected costs by taking the precaution for which the liability marginal cost is equal to the effect minus the injury effect.

Whether a little more or less precaution is better at minimizing the firm's costs depends on the marginal cost of precaution and on the relative size of the liability and injury effects. We cannot be certain of this without making some additional assumptions about the components of $TC(x)$.

For present purposes, the important assumptions concern the nature of the distribution q(x) and the behavior of the marginal cost function C′(x).

Turning first to assumptions concerning the nature of the distribution q(x), we can show (see Johnson *et al.*) the following two propositions.

Proposition 1. Assuming a mean-preserving spread, an increase (decrease) in the uncertainty surrounding the legal standard will result in firms' reducing (increasing) their level of precaution. If the uncertainty in determining causality is great (small) enough, firms will take too little (much) precaution *vis-á-vis* the social optimum.

This case is shown in Figure 2.[11] This figure demonstrates that even if the firm's expected value of precaution is x^* (the social optimum) the presence of a sufficient degree of uncertainty will result in under- or over-precaution.

Proposition 2. If the firm's subjective distribution is biased so that its mean is below (above) the social optimum, then the firm also will take too little (much) precaution.

The last case is shown in Figure 3. Note that as the distribution shifts to the right (left) the variance of the distribution increases (decreases). This need not be the case. Examination of Proposition 1 suggests that if the shift in the distribution preserved variance, then Proposition 2 would still hold. The result of Proposition 2 is intuitively obvious and is not as insightful as that for Proposition 1. ***It should be made clear that neither of the above two propositions nor those that follow depend on the firm exhibiting risk-averse behavior. The firm is assumed to be risk neutral.***

---

[11] The mathematical formulation of the appropriate density function presented in this figure is from Johnson *et al*. The density function is a particular type of mean-preserving spread. It is used for a similar purpose later in this paper to derive a proposition with regard to uncertainty of enforcement of an *ex ante* regulation.

Regarding assumptions concerning the behavior of the marginal cost function C′(x), we previously developed a third proposition regarding whether the firm would over- or under-protect against the environmental harm:

Proposition 3. Given a socially optimal level of precaution greater than zero, and a sufficiently large (small) marginal cost of precaution at the social optimum, x^*, then the firm will employ too little (too much) precaution when facing strict liability.

Like Proposition 2 this proposition is intuitively obvious. The results stated in Propositions 1-3 are similar to those of Craswell and Calfee 1984 and 1986), who analyzed the behavior of a firm facing a negligence rule with uncertainty regarding the legal standard.

Earlier, we assumed that victims were perfectly compensated and the focus was on uncertainty regarding the determination of the proximate cause of the harm. Now we turn to the case where uncertainty regarding proximate cause is fixed but compensation is imperfect. The effect of imperfect compensation on the firm's precautionary decision can be stated as a fourth proposition:

Proposition 4. Assume that the firm's expectation of the costs of a harm is greater (less) than A(x), society's expectation regarding these costs. If so, then the firm will oversupply (undersupply) precaution when subject to strict liability.

This proposition is illustrated in Figure 4.

An example of a potential case in the environmental regulatory arena to which the above model and propositions might apply is that of the small-firm generator of hazardous wastes. Prior to the 1984 Amendments to the Resource Conservation and Recovery Act (ARCRA), small-firm generators of hazardous wastes faced little or no regulatory requirements in the handling of such wastes. These generators did however face public liability actions under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1982. But given the lack of record keeping regarding their disposal of hazardous wastes and therefore the difficulty of establishing causality, and problems with USEPA's Superfund Program there was a great deal of uncertainty regarding enforcement. Along with this great uncertainty were the relatively high marginal costs to these small firms of proper hazardous waste disposal and the reality that only public causes of action held any real chance of successful litigation. Given these factors, Propositions 1,3, and 4 would point to a high degree of under-precaution by these firms. Congress may have intuitively understood this when the small generator provisions in ARCRA were made considerably tighter.

B. A Model of the Firm Facing an *Ex Ante* Regulation

We now abandon strict liability for the moment to examine the case of the firm's behavior when facing an *ex ante* regulation. Because the deterministic model of the firm under these circumstances is well known, we will proceed immediately to the case of uncertain enforcement. Our task is greatly simplified by noting that the analysis of an uncertain *ex ante* regulation looks very much like that of an uncertainly enforced negligence rule, as described in Johnson *et al.* and Craswell and Calfee (1984 and 1986).

For the case of a fixed fine for violating the regulation the firm's objective function can be formulated as

$$\min_{x} \text{Total Costs} = TC(x) = C(x) + F(y)p(x) \qquad (6)$$

where C(x) is defined as the costs of precautionary measures to meet the regulation, F(y) is a fixed fine for not meeting the required level of precaution (y) ($F(y) > 0$ for $x < y$ and $F(y) = 0$ for $x \geq y$), and P(x) is the probability of being found in violation of the regulation. P(x) like R(x) in our earlier model of strict liability is given by

$$P(x) = \int_{x}^{\infty} g(x)dx \qquad (7)$$

where g(x) is a continuous probability density with a support $[0,\infty)$. as before we will assume that $[C(X) + F(_y)P(X)]$ is strictly convex. We will also asume that y is set equal to the social optimum. The first-order condition for the minimization of Equation 6 is

$$TC'(\bar{x}) = C'(\bar{x}) - F_{(}y)g(\bar{x}) = 0 \qquad (8)$$

or

$$C'(\bar{x}) = F(_y)g(\bar{x}).^{12} \qquad (9)$$

An obvious question is how the distribution g(x) and the marginal cost function C'(x) will affect the choice of $\bar{x}$ relative to the regulatory standard y. That is, can we tell whether uncertain enforcement of the regulatory standard punished by a fixed fine will induce optimal precaution, over-precaution, or under-precaution? We can determine this by evaluating

---

[12] The exposition here closely follows that of Craswell and Calfee and Johnson *et al.* for a firm facing a negligence rule with a fixed fine.

the sign of $TC'(\tilde{x})$ in Equation 8 when the firm's precaution is just equal to the regulatory standard. Because $TC(x)$ is strictly convex and is minimized at $\tilde{x}$, $TC'(y) < 0$ when $y < \tilde{x}$ and $TC'(y) > 0$ for $\tilde{x} < y$.

As in our discussion of the firm's behavior when facing a strict liability standard, let the exposition first turn to assumptions regarding $g(x)$. To facilitate our comparative static analysis, we introduce a particular type of mean-preserving spread on $g(x)$:

$$g_\gamma(x) = \begin{cases} \gamma g_\gamma[a(x-y) + y,], & \text{for } x \geq \frac{y(\gamma-1)}{\gamma} \\ 0 & \text{, otherwise} \end{cases} \tag{10}$$

where y is the expected value of x when x is distributed as $g(x)$. It is easily seen that $g_1(x) = g(x)$. The density function $g_\gamma(x)$ is well-behaved for all values of $\gamma > 0$, and random variables distributed according to g and g_γ have the same mean. A decrease in γ causes the spread of g_γ to increase, and an increase in γ causes the probability mass to become concentrated at the mean similar to what occurs in Figure 2 for $q_a(x)$, a hypothetical distribution.

As uncertainty concerning enforcement of the *ex ante* regulation becomes greater (less) for the firm, i.e., as γ gets smaller (larger), $g(y)$ becomes smaller (larger) causing $TC'(y)$ in Equation 8 to become positive (negative). This implies that $\tilde{x}$ is less (greater) than y. This result can be made into a formal proposition.

Proposition 5. Assume that uncertainty regarding the enforcement of the regulatory standard y is distributed according to $g(x)$ with $g(x) > 0$. If (in the sense of Equation 10) uncertainty is sufficiently large (small), then the firm subject to the standard will take too little (much) precaution.

The proof of this follows that for proving Proposition 1 given in Johnson *et al*. It can be shown that a sufficiently spread out (compact) version of g(x) exists such that TC'(y) in Equation 8 becomes positive (negative). Since C'(x) > 0, there exists a $\gamma > 0$ such that TC'(y) becomes positive and at least one value such that TC'(y) becomes negative.

Having the firm's expected value of precaution equal to the regulatory standard is not sufficient to insure that the level of precaution employed by the firm is in fact equal to that standard. This is true even for come risk-neutral firms. (Note the analogy between Propositions 1 and 5.)

This last proposition should convince the reader of the equivalency of an analysis of uncertain enforcement of a regulatory standard and that of uncertain enforcement of a negligence or strict liability rule. This will allow us to expedite the formulation of two more propositions regarding uncertain enforcement of a regulatory standard. The first of these regards biased perceptions of g(x):

Proposition 6. If the distribution g(x) is sufficiently biased in the sense of Figure 2 to the left (right) of y, and q(y) > 0, then the firm will take too little (little) precaution.

The proof of this proposition follows that given Johnson *et al*. for proving Proposition 2. While not proving a similar result, Craswell and Calfee show by means of simulation that the exact assumptions regarding the bias for either a shift in the mean or change to a non-symmetric distribution around the standard do not affect Proposition 6.

Next we will deal with the effect of marginal cost on the firm's choice of $\tilde{x}$ when there is uncertain enforcement of the regulatory standard.

Proposition 7. Given $g(y) > 0$, then for suitably large (small) marginal cost of precaution at the regulatory standard, y, the firm will employ too little (much) precaution *vis-á-vis* the standard.

(Note that Proposition 7 is equivalent to Proposition 3 presented earlier.)

By now the reader should be aware that Propositions 5 through 7 would hold if the firm faced a fine based on the level of precaution. The firm's objective function could be written as

$$\min_{x} TC(x) = C(x) + F(y-x)P(x) \tag{11}$$

where $F(y-x)$ is the fine function that meets the following conditions: $F(\cdot) \geq 0$ $(\chi \geq 0)$, equals zero when $x \geq y$, and $\partial F(\cdot)/\partial x < 0$.

We may summarize the results of this section in the following way: if any or all of the following conditions are true——there is great (small) uncertainty regarding enforcement of the regulation; there is a leftward (rightward) biased perception of the uncertainty surrounding enforcement of the regulation; or there are large (small) marginal costs at the regulatory standard, then Propositions 5 through 7 state that we would expect the firm to take too little precaution by comparison to the regulatory standard.

C. A Model of the Firm Facing Both *Ex Ante* Regulation and *Ex Post* Strict Liability

We now combine the separate analyses of uncertain enforcement of *ex ante* regulation and *ex post* liability in order to see under what conditions this simultaneous use of *ex ante* and *ex post* policies is efficient. We will first look at the case of uncertain enforcement of strict liability but certain enforcement of a regulatory standard.[13]

Introduction of the *ex ante* regulatory standard can be made by assuming that the firm's distribution around the social optimum is truncated. We assume that, given a regulatory standard, y, the firm will not take precautionary measures below y and that, therefore, the probability distribution around the social optimum is truncated. The result of this assumption is to move the probability mass that would be to the left of y without a regulatory standard to the right of y, i.e., there is a zero probability that the firm's choice of x will be below y. This will have the effect of raising the density function to the right of y and shifting the mean of the resulting probability distribution. The intuition behind this truncation assumption is that after the imposition of a regulatory standard, the firm perceives that the further its choice of x is to the right of y the less likely it will be found liable. This truncation assumption allows us to model the firm's behavior without a fine being associated with violation of the *ex ante* regulatory standard. Truncation alone is sufficient to affect the firm's behavior.

The new density function, $\underline{q}(x)$, that results from our truncation assumption has a conditional distribution

$$\underline{q}(x) = q(x \mid x \geq y). \qquad (12)$$

The conditional probability $\underline{R}(x)$ that the firm will be found strictly liable if its level of precaution is x is

$$\underline{R}(x) = \frac{R(x)}{R(y)}. \qquad (13)$$

---

[13] A certain regulatory standard can be thought of as having a point distribution regarding uncertainty of enforcement.

The firm's objective function becomes

$$\min_{x} TC(x) = C(x) + A(x)\underline{R}(x). \tag{14}$$

The first-order condition for this minimization is

$$TC'(\tilde{x}) = R(y)C'(\tilde{x}) + A'(\tilde{x})R(\tilde{x}) - A(\tilde{x})q(\tilde{x}), \tag{15}$$

were $\tilde{x}$ is now understood to mean $\tilde{x}(y)$.

To show the firm's response to changes in the regulatory standard requires total differentiation of Equation 15 and the evaluation of the sign of $d\tilde{x}/dy$. The result of the total differentiation is

$$\frac{d\tilde{x}}{dy} = \frac{q(y)C'(\tilde{x})}{R(y)C''(\tilde{x}) + A''(\tilde{x})R(\tilde{x}) - 2A''(\tilde{x})q(\tilde{x}) - A(\tilde{x})q'(\tilde{x})} \tag{16}$$

Since $[C(x) + A(x)\underline{R}(x)]$ is convex, the denominator of Equation 16 is positive. The numerator is also positive, which implies that $d\tilde{x}/dy$ is greater than zero. Therefore, increasing the regulatory standard, y, has the impact of increasing the level of precaution taken. The results of this analysis can be stated as a proposition (Johnson *et al.*).

Proposition 8. Given that the firm faces strict liability with uncertain enforcement, imposition of a certain *ex ante* regulatory standard will always induce more precaution by the firm. This will promote efficiency if the firm would under-protect without *ex ante* regulation and will exacerbate inefficiency if the firm would over-protect in the absence of *ex ante* regulation.

In principle the optimal level of the regulatory standard, y^*, needed to induce the firm to choose x^* can be determined. However, the relationship between y^* and x^* can be shown by simply rewriting Equation 15 after substituting in Equation 2, as

$$C'(x^*)[R(y^*) - R(x^*)] - A(x^*)q(x^*) = 0. \tag{17}$$

Evaluation of this equation gives us another proposition.

Proposition 9. The optimal *ex ante* regulatory standard, y^*, given its certain enforcement and the uncertain enforcement of *ex post* liability, will be less than the socially optimal level of precaution, x^*, provided $q(x^*) > 0$. $y^* = x^*$ iff $q(x^*) = 0$.

The situation of a hazardous waste landfill firm facing both ARCRA and CERCLA can now be evaluated in light of Propositions 8 and 9. Under CERCLA the firm faces potential pre- and post-closure liability with uncertain enforcement for the cleanup of leaks from the site. If the landfill site is currently not a Superfund site, then the uncertainty regarding pre- and post-closure liability may be great. This uncertainty, coupled with the fact that in all likelihood the firm will only face public causes of action and that as a result there will be less than perfect compensation of victims, will result in the firm's providing too little precaution. However, under ARCRA the firm faces nearly certain enforcement of ARCRA regulations regarding the construction of the landfill site, and the handling and recording of wastes received. These *ex ante* regulations should force the firm to chose a higher level of precaution than it would under CERCLA alone and thus to move toward the social optimum. It is still an open question as to whether the regulations are sufficient to shift the firm's choice of level of precaution to the social optimum.

The final embellishment to the present model is to include the possibility that there is uncertain enforcement of both the *ex ante* regulatory standard and *ex post* liability. In adding uncertainty regarding the enforcement of *ex ante* regulation, we must look at the possible states of the world that the firm faces.

1. The firm will be found in violation of the regulatory standard and therefore found liable for all environmental harms stemming from such a violation.
2. The firm will not be found in violation of the regulatory standard but will be found strictly liable for the environmental harms.
3. The firm will not be found in violation of the regulatory standard nor will it be held strictly liable for the environmental harms.

The first two of these outcomes indicate a conditional relationship between being found in violation of the regulatory standard and being held strictly liable for alleged environmental harms. We can think of the firm facing a sequential process--it is first determined whether or not it has violated the regulatory standard; then it is determined if it is strictly liable for environmental harms. If the firm is not in violation of the regulatory standard, then in the second step of the process it is assumed that its level of precaution is at least equal to the standard. This allows the retention of $\underline{q}(x)$ as the appropriate density function.

Given a fixed fine for violation of the regulatory standard, the firm's objective function can now be written as

$$\min_{x} TC(x) = C(x) + F(y)P(x) + A(x)\{P(x) + [1-P(x)]\underline{R}(x)\}. \tag{18}$$

As before we will assume that the right-hand-side of Equation 18 is strictly convex. The first-order condition for Equation 18 is

$$\begin{aligned} TC'(\tilde{x}) = {} & R(y)\{C'(\tilde{x}) + A'(\tilde{x})P(\tilde{x}) - g(\tilde{x})[F(y) + A(\tilde{x})]\} \\ & + R(\tilde{x})\{A'(\tilde{x})[1 - P(\tilde{x})] + g(\tilde{x})A(\tilde{x})\} \\ & + q(\tilde{x})P(\tilde{x})A(\tilde{x}) = 0. \end{aligned} \tag{19}$$

As was the case above we need to totally differentiate Equation 19 and determine the sign of $d\bar{x}/dy$. The results of total differentiation, upon rearrangement of the terms, is

$$\frac{d\bar{x}}{dy} = \frac{q(y)(C'(\bar{x}) + A'(\bar{x})P(\bar{x}) - g(\bar{x})[F(y) + A(\bar{x})])}{D(\bar{x},y)} \tag{20}$$

where

$$\begin{aligned} D(\bar{x},y) = {} & R(y)(C''(\bar{x}) + A''(\bar{x})P(\bar{x}) - 2g(\bar{x})A'(\bar{x}) - g'(\bar{x})[F(y) + A(\bar{x})]) \\ & + A''(\bar{x})R(\bar{x})[1-P(\bar{x})] - A'(\bar{x})g(\bar{x})[1 - P(\bar{x})] + 2A'(\bar{x})R(x)g(x) \\ & - 2A(\bar{x})g(\bar{x})q(\bar{x}) + [P(\bar{x}) - 1][q(\bar{x}) + A(\bar{x})q'(\bar{x})] + q'(\bar{x})A(\bar{x})R(\bar{x}), \end{aligned} \tag{21}$$

the second derivative of the total cost function $TC(x)$.

$D(\bar{x},y)$ is positive because Equation 18 was assumed to be strictly convex. The numerator of Equation 19 is also positive if the uncertainty surrounding the enforcement of the regulatory standard is in some sense small, which implies that $d\bar{x}/dy$ is greater than zero.[14] Thus in this case, just as was the case for the regulatory standard given certain enforcement, increasing the minimally acceptable regulatory standard has the effect of increasing the precaution taken. This outcome follows from the fact that the determination of violation comes prior to the judgement of liability. Note that if $d\bar{x}/dy \quad 0$, this would be a *prima facie* for the abandonment of strict liability because y^* would need to be set greater than or equal to the social optimum, x^*.

The implication of the above findings is that if $\bar{x} < x^*$ (i.e., $\bar{x}(0) < x^*$) but greater than y prior to the imposition of the *ex ante* regulation, then the introduction of regulation will promote efficiency. If, on the contrary, $\bar{x}(0) > x^*$, then *ex ante* regulation with small uncertainty will exacerbate the inefficiency that exists when there is uncertain enforcement of a strict liability rule.

---

[14] The definition of small uncertainty here is that which will result in too much precaution regarding the regulatory standard, i. e., $\bar{x} > y$. If this is the case the term $F(y)$ equals zero.

Proposition 10. The imposition of an *ex ante* regulatory standard, given the existence of a strict liability rule, will promote efficiency if the firm would under-protect in the absence of the standard and will exacerbate inefficiency if the firm would over-protect without the standard. This only holds if the choice of $\tilde{x}$ is greater than y implying that there is relatively little uncertainty regarding the enforcement of the regulatory standard.

Finally we can explore the relationship between the optimal level of the *ex ante* regulatory standard, y^*, and the socially optimal level of precaution, x^*. From Proposition 10 we know that $y^* = 0$ if and only if $\tilde{x}(0) \geq x^*$. For $\tilde{x} < x^*$, $y^* > 0$ will promote efficiency. What level of y will make $\tilde{x} = x^*$? The answer can be found by substituting x^* for $\tilde{x}$ in Equation 19 and solving for y^*. However, While the actual computation of y^* can be computed, the regulators must know the cost of precaution for the firm and the characteristics of the density functions q(x) and g(x). Stopping short of actually solving Equation 19 for y^*, we can rewrite it, using Equation 2, as

$$\begin{aligned} TC'(\tilde{x}) = {} & [R(y^*) - R(x^*)][C'(x^*) - C'(x^*)P(x^*)] \\ & - R(y^*)g(x^*)A(x^*) + R(x^*)g(x^*)A(x^*) \\ & + q(x^*)P(x^*)A(x^*) = 0.^{15} \end{aligned} \tag{22}$$

For the equality with zero to hold, the term $[R(y^*) - R(x^*)]$ must be positive. This implies that the optimum level of the regulation is less than or equal to the optimal level of precaution, i.e., $y^* \leq x^*$.

---

[15] Note that the term F(y) is equal to zero at x^* and does not appear in Equation 22.

Proposition 11. The optimal level of the *ex ante* regulatory standard, y^*, given that strict liability exists, will be less than the socially optimal level of precaution, x^*, provided $q(x^*) > 0$ and $g(x^*) > 0$.

Note that unlike the case of a certain *ex ante* regulatory standard There will be no case where $y^* = x^*$, because there is uncertainty surrounding the enforcement of y^*. It should also be intuitive that with uncertainty surrounding the enforcement of y^* being small in the sense used in Proposition 5, the distance between y^* and x^* will be greater with uncertain enforcement than with certain enforcement.

The implication of the above finding is that where the optimal precaution calls for the joint use of *ex ante* regulation and *ex post* strict liability the regulatory standard should be set below the socially optimal level of precaution.

IV. Conclusions

There are two groups of conclusions that can be drawn from the above research. The first group of conclusions is regarding the use of either an *ex ante* regulatory standard or *ex post* strict liability rule by themselves when either has uncertain enforcement. The second group of conclusions concerns the complementary use of *ex ante* regulation and *ex post* strict liability.

The first conclusion regarding the sole use of either *ex ante* regulation or *ex post* strict liability when neither has certain enforcement is that if the uncertainty has an expectation of the social optimum, x^* (or $y^* = x^*$), neither policy will result in an efficient outcome from society's point of view. With regard to either policy the firm will take to little (much) precaution if:

1. according to Propositions 1 or 5, depending on the policy, if firm's uncertainty regarding enforcement is too large (small);

2. according to Propositions 2 or 6, depending on the policy, if the firm's subjective distribution regarding uncertainty of enforcement is sufficiently biased below (above) the social optimum; and

3. according to Propositions 3 or 7, depending on the policy, if the firm's marginal costs at the socially optimal level of enforcement is large (small).

Furthermore, in the case of *ex post* strict liability, according to Proposition 4 under-(over-)estimation of victim compensation will result in too little (much) precaution. Given the great deal of uncertainty surrounding the monitoring and enforcement of *ex ante* regulation and the uncertainty surrounding the enforcement of *ex post* strict liability with regard to environmental problems, it appears that the sole reliance on either one policy or the other is unwise.

We can also derive important conclusions regarding the complimentary use of both *ex ante* regulation and *ex post* strict liability. First, when the joint use of these two policies is desirable then the optimal regulatory standard, y^*, should be less than the social optimum. Also, if there is small uncertainty in the sense of Proposition 5 surrounding the use of an *ex ante* regulatory standard, the optimal level of y^* should be set lower than that if there was certain enforcement. Finally, the *ex ante* regulatory standard should be used by itself if there is certain enforcement and the firm's density function, $q(x)$, regarding the enforcement of the strict liability rule equals zero; or a movement of the *ex ante* regulatory standard toward x^* does not result in a higher level of precaution on the part of the firm. This may point to the correctness of regulating hazardous wastes with both ARCRA and CERCLA.

Finally, the validity of the above modeling effort can only be proved regarding the real world by empirical work. The most suitable method for empirical analysis may be simulation as employed by Craswell and Calfee (1986).

Bibliography

Cooter, Robert D. and Thomas S. Ulen, "An Economic Case for Comparative Negligence," *New York University Law Review*, 1986.

Cooter, Robert D. and Thomas S. Ulen, *Law and Economics*, Glenview, IL: Scott, Foresman and Co, 1987.

Craswell, R. and John E. Calfee, "Some Effects of Uncertainty on Complience with Legal Standards," *Virgina Law Review*, 1984, 70.

Craswell, R. and John E. Calfee, "Deterrence and Uncertain Legal Standards," *Journal of Law, Economics, and Organization*, Fall 1986, 2(2).

Harvard Law Review, "Developments in the Law—Toxic Waste Litigation," May 1986, *99*.

Johnson, Gary V., Charles D. Kolstad, and Thomas S. Ulen, "*Ex Ante* Regulation and *Ex Post* Liability: Substitutes or Complements?" (unpublished ms. 1987).

Johnson, Gary V. and Thomas S. Ulen, "Designing Public Policy Toward Hazardous Wastes: The Role of Administrative Regulations and Legal Liability Rules," *American Journal of Agricultural Economics*, 1986, *68*.

Russell, Clifford S., Winston Harrington, and William J. Vaughan, *Enforcing Pollution Control Laws* (Washington, D.C.: Resources for the Future, 1986).

LEGENDS

Figure 1.
The Social Problem with Evidentuary Uncertainty for the Injurer

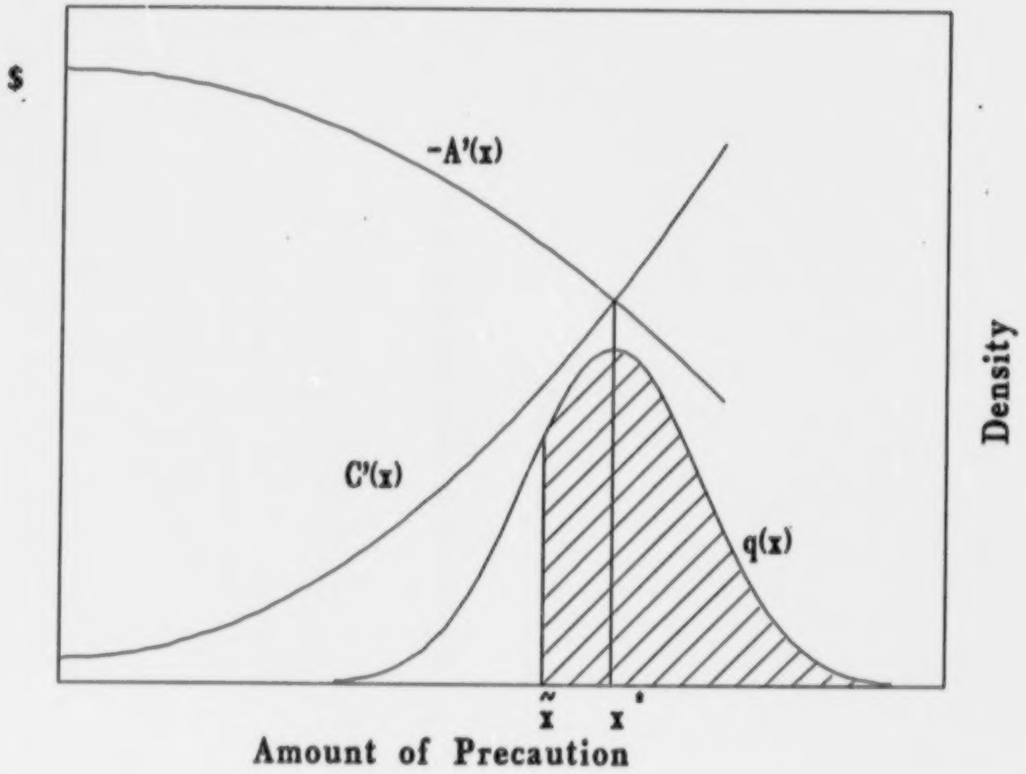

$
-A'(x)
C'(x)
q(x)
Density
$\tilde{x}$
x^*
Amount of Precaution

Figure 2.
Two Levels of Uncertainty Around x^*

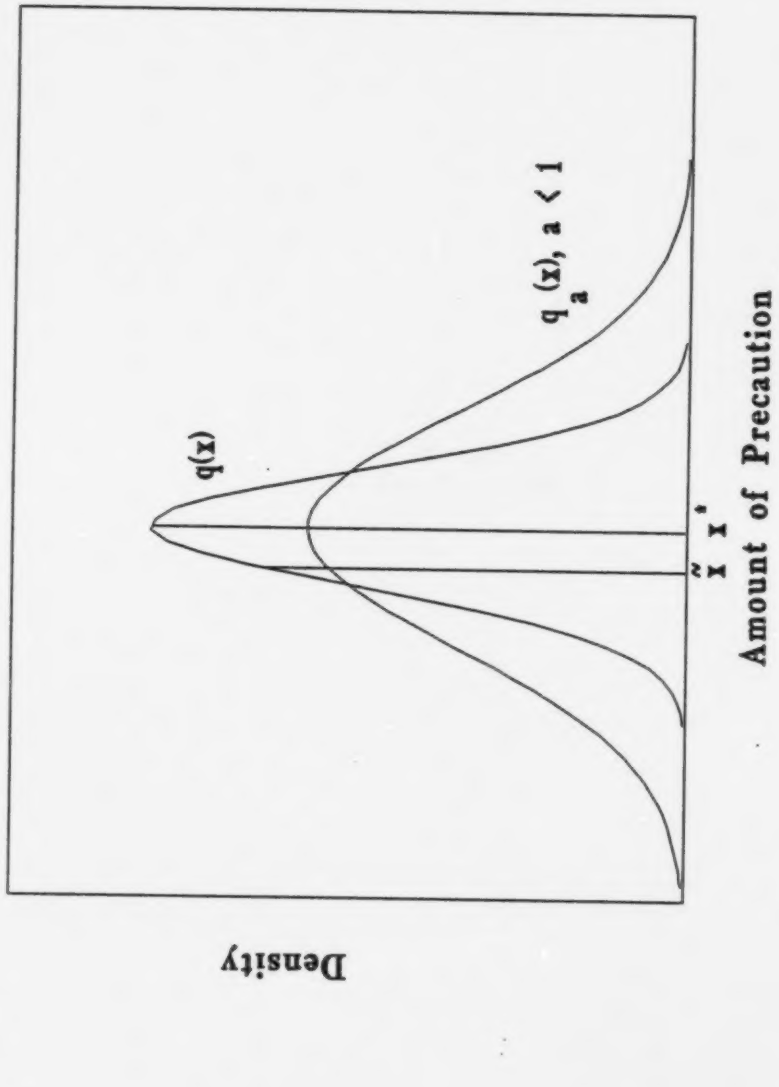
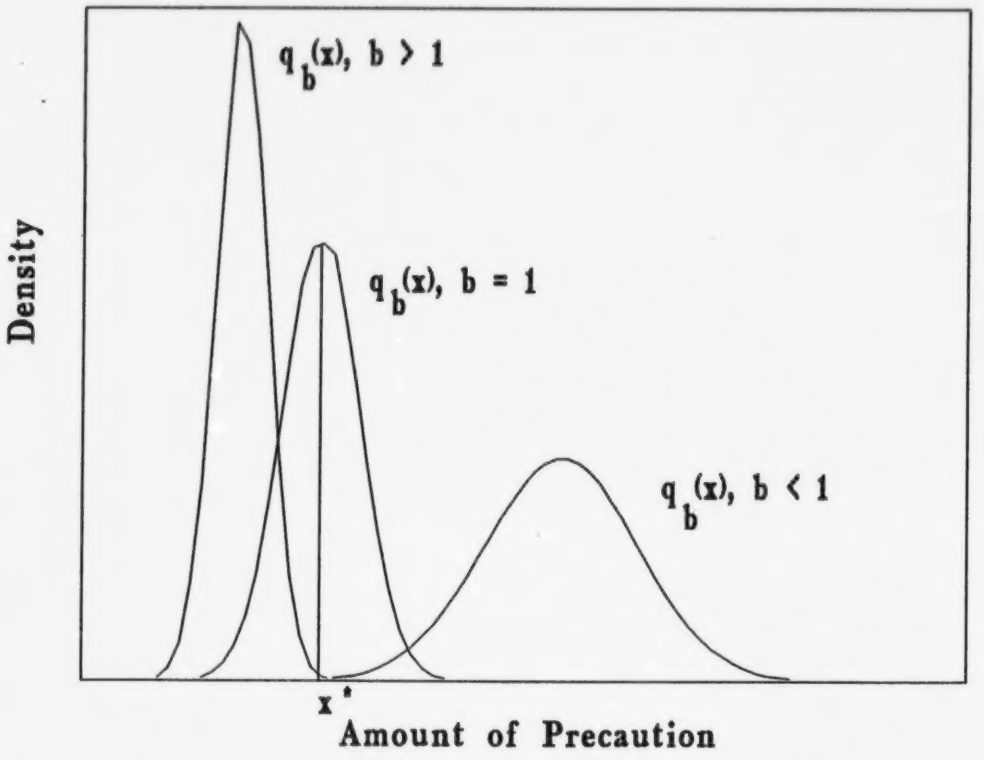
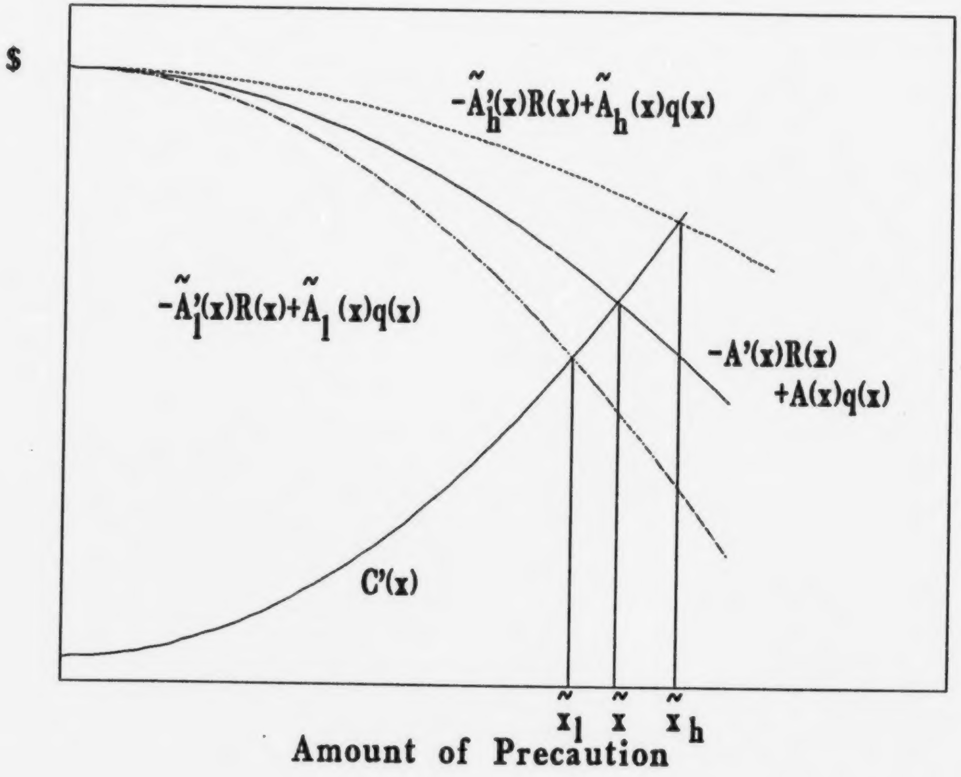

q(x)
q_a(x), a < 1
x̃
x*
Amount of Precaution
Density

Figure 3.
Biased Uncertainty



$q_b(x)$, b > 1
$q_b(x)$, b = 1
$q_b(x)$, b < 1
Density
x^*
Amount of Precaution

Figure 4.
The Injurer's Decision With Imperfect Compensation


$
$-\tilde{A}_h'(x)R(x)+\tilde{A}_h(x)q(x)$
$-\tilde{A}_l'(x)R(x)+\tilde{A}_l(x)q(x)$
$-A'(x)R(x)$
$+A(x)q(x)$
$C'(x)$
$\tilde{x}_l$
$\tilde{x}$
$\tilde{x}_h$
Amount of Precaution

MARKETS FOR POLLUTION CONTROL WHEN FIRMS ARE NONCOMPLIANT

Arun S. Malik

Bureau of Business and Economic Research
University of Maryland
College Park, Maryland 20742
July 8, 1987

1. Introduction

Over the past few years considerable support has developed for establishing transferable discharge permit (TDP) markets for controlling pollution. This interest in TDP markets is the result of a growing disenchantment with existing policies that rely on direct controls in the form of technology-based emissions standards. In addition to criticizing existing policies on the grounds of allocative inefficiency, many analysts have emphasized the problems of enforcing them in the face of widespread noncompliance.[1] Little attention has been given, however, to the consequences of noncompliance for the alternative policy instruments prescribed such as TDP markets.

The problems encountered in enforcing the current policy have been due largely to the technical difficulties in continuously monitoring pollutant emissions, and the absence of well-developed mechanisms for assessing penalties for noncompliance. As long as these deficiencies persist, noncompliance and the consequent need for enforcement will be an intrinsic part of any effective pollution control policy.

This paper examines the consequences of noncompliance for a TDP market. It does not attempt to characterize the optimal policy for enforcing such a market. Rather, it examines the effects of noncompliance given fairly general assumptions about the characteristics of the enforcement policy in place. This is accomplished by formally incorporating noncompliant behavior in a standard market model.

The first part of the paper examines questions such as whether TDP markets retain their efficiency properties in the presence of noncompliance, and what the effects of noncompliance are on the equilibrium permit price. The second part of the paper compares the consequences of noncompliance for a TDP market with those for an effluent tax policy, and discusses the problems that are likely to be encountered when implementing a TDP market given noncompliant behavior.

The existing literature on noncompliance and the enforcement of environmental policies is quite small. The first theoretical work is due to Harford (1978), who examines the behavior of a noncompliant, risk-neutral firm under two different policy instruments -- standards and taxes. A similar analysis is presented by Storey and McCabe (1980) for the case of a risk-averse firm. The only formal treatment of noncompliance in a TDP market is contained in a paper by Beavis and Walker (1983). They develop a simple model of a TDP market consisting of risk-neutral firms with stochastic pollutant discharge. The primary aim of their paper is to demonstrate that the frequency of monitoring does influence pollutant discharge when the latter is stochastic, contrary to what is allegedly found by Harford under the assumption that discharge is deterministic. As is shown below, Beavis and Walker incorrectly apply Harford's effluent tax model to a TDP market. Pollutant discharge is, in fact, a function of the frequency of monitoring whether discharge is stochastic or not.

A simplifying assumption made in this paper is that a firm's location is irrelevant: a unit of pollutant discharged by one firm is identical to that of any other firm insofar as its effect on ambient quality. Achieving a total discharge goal is then equivalent to maintaining a prescribed level of ambient quality, the usual context in which TDP markets are discussed. This assumption merely allows for an economy of notation; relaxing it does not materially alter the results presented.

2. The Model

The TDP market consists of n firms and a central authority responsible for issuing permits and ensuring that firms do not discharge more than the permitted amount. A total of s_T permits are issued by the central authority. These are allocated among the n firms via an auction or some other suitable mechanism. The authority may, for instance, initially distribute permits free, based on some equity criterion, and then allow firms to trade them. The firms are assumed to be price takers in the permit market and in their output markets.

The ith firm's profits from engaging in the pollutant generating activity, excluding permit and fine payments, are given by a smooth, strictly concave function $B_i(w_i)$ of pollutant discharge.[2] As is plausible, profits first increase in w_i, reaching a maximum at w_i^o, and then decline, as it becomes necessary for the firm to devote resources to generating pollution to further raise its discharge level. Clearly, the firm would not operate in the latter region.

The n firms face a common permit price r. Permit holdings are denoted by s_i, and the number of permits a firm initially receives free, if any, by s_i^o. If a firm is compliant, $w_i \le s_i$ and the magnitude of its violation, v_i, is zero. If a firm is noncompliant, $w_i > s_i$, and the magnitude of its violation is given by $v_i = w_i - s_i$.

Audit Probability and Penalty for Violations

The exact amount of pollutant discharged by a firm can be determined by the central authority only by conducting a compliance audit. It is assumed that these audits are unannounced and that firms cannot vary their permit holdings or pollutant discharge levels during an audit.

The central authority's decision on whether or not to audit a firm may depend on a variety of factors. These include the number of permits held by the firm (which is presumably public knowledge) and an estimate of the firm's pollutant discharge. The authority's ability to derive such an estimate, and its accuracy, will depend on the nature of the pollutant, its effects on environmental quality, and the characteristics of the firm's production process. For instance, particulate emissions from a smokestack can be estimated quite easily by remote monitoring. However, estimating discharges of water pollutants is likely to be difficult without conducting an on-site audit.

From the firm's perspective, audits are uncertain. This may be due to intentional randomness in the authority's audit policy or to uncertainty on the firm's part regarding the authority's estimate of pollutant discharges. Thus, a firm's *subjective* probability of being audited is given, quite generally, by $p_i(w_i,s_i;\theta)$, where θ is a vector of audit policy parameters set by the central authority. Since the enforcement policy is taken as given, this parameter vector generally will be omitted. Note that the subscript on $p_i(.)$ implies that the audit probability functions are allowed to vary across firms. The presence of w_i in the probability function reflects the assumption that the authority's estimate of pollutant discharges is a function of actual discharges.[3] Assuming the estimate is an increasing function of actual discharges, we would expect the audit probability to be non-decreasing in w_i. The relationship between the level of permit holdings and the audit probability is more difficult to specify. It can be argued that the probability is increasing in s_i over some intervals and decreasing over others. I allow for both these possibilities.

It should be noted that the specification of the audit probability function includes the special cases of a constant audit probability, that is, one independent of w_i and s_i, and an audit probability dependent only on the firm's violation size, $v_i = w_i - s_i$.

If audited and found in noncompliance, the firm incurs a certain penalty, the magnitude of which is given by a smooth, increasing function of the firm's violation size $F_i(v_i;\delta)$, where δ is a vector of penalty policy parameters set by the enforcement agency. As with the audit parameters, the penalty parameters generally will be omitted below.

Firm's Decision Problem

Given the audit uncertainty, each firm seeks to maximize its expected utility of profits, $E[U_i(\pi_i)]$,

$$(1) \qquad \max_{w,s} \; [(1-p_i)U_i(\pi_i^0) + p_iU_i(\pi_i^1)]$$

where

$$\pi_i^0 = B_i(w_i) - r(s_i - s_i^0)$$
$$\pi_i^1 = \pi_i^0 - F_i(v_i)$$

and $p_i = p_i(w_i,s_i)$.

3. Optimal Pollutant Discharges and Permit Holdings

Pollutant Discharge

Omitting the firm subscript, the first-order conditions for an interior solution to (1) are:

$$(2) \qquad (1-p)U'(\pi^0)B' + pU'(\pi^1)(B' - F') \;=\; \frac{\partial p}{\partial w}[U(\pi^0) - U(\pi^1)]$$

(3) $$-(1-p)U'(\pi^0)r + pU'(\pi^1)(-r + F') = \frac{\partial p}{\partial s}[U(\pi^0) - U(\pi^1)]$$

Adding (2) and (3) and rearranging terms yields

(4) $$B_i'(w_i) = r + \phi_i\left[\frac{U_i(\pi_i^0) - U_i(\pi_i^1)}{E[U_i'(\pi_i)]}\right]$$

where $\phi_i = \partial p_i/\partial w_i + \partial p_i/\partial s_i$. The key expression in this equation is the second term on the RHS. Since $[U(\pi^0)-U(\pi^1)]/E[U'(\pi)] > 0$, the sign of the second term depends on that of ϕ_i. As noted earlier, $\partial p_i/\partial w_i$ is likely to be non-negative, but $\partial p_i/\partial s_i$ may be positive or negative. Hence, at this level of generality, the sign of ϕ_i is ambiguous. Note, however, that if

(5) $$\phi_i = \partial p_i/\partial w_i + \partial p_i/\partial s_i = 0$$

over the relevant range, (4) reduces to

(6) $$B_i'(w_i) = r,$$

which is identical to the abatement decision rule employed by a firm that is, by assumption, perfectly compliant. (A compliant firm maximizes $B_i(w_i) - r(w_i-s^0)$.)

Thus, if $\phi_i = 0$, the quantity of pollutant discharged by a noncompliant firm for a given permit price is equal to that discharged by an otherwise identical compliant firm: $w_i^*(r) = \overline{w}_i^*(r)$, where $\overline{w}^*$ denotes the pollutant discharge of the compliant firm. An obvious corollary to this result is that when $\phi_i = 0$ the quantity of pollutant discharged by a noncompliant firm does not directly depend on the enforcement policy, or on the firm's attitude towards risk.

Although the condition $\phi_i = 0$ is fairly stringent, there are two interesting cases when it does hold. The first is when, over the relevant range, the subjective audit probability is a constant, independent of the firm's decisions, then, $\partial p_i/\partial w_i = \partial p_i/\partial s_i = 0$. The second case is when, over the relevant range, the audit probability is a function of the firm's violation size, that is, $p_i(w_i,s_i) = P_i(w_i - s_i)$; then, $\partial p_i/\partial w_i = P_i'$ and $\partial p_i/\partial s_i = -P_i'$. For these two cases, the abatement decision rule employed by a noncompliant firm (6) is identical to that used by a compliant firm.

In general, however, the presence of the second term on the RHS of (4) drives a wedge between the permit price (r) and marginal profits (B_i'). The sign of this term depends on the characteristics of the audit probability function. Its magnitude is determined, in part, by the firm's attitude towards risk. The strict concavity of the profit function implies that if the second term is negative, the quantity of pollutant discharged by a noncompliant firm is larger than that of an otherwise identical compliant firm ($w_i^* > \overline{w}_i^*$). Conversely, if the second term is positive, the quantity of pollutant discharged is smaller ($w_i^* < \overline{w}_i^*$).

Permit Demand and Equilibrium Permit Price

The above results present somewhat of an enigma. They indicate that, for a given permit price, the quantity of pollutant discharged by noncompliant firms may be equal to or even smaller than the quantity discharged by otherwise identical compliant firms. Yet, by definition, aggregate discharge must be higher in a TDP market in which firms are noncompliant. Upon reflection, it is apparent that for these two observations to be reconciled, the equilibrium permit price in a noncompliant TDP market must differ from that in a compliant market.

Examining the first-order condition for s_i (3), it is clear that a noncompliant firm's demand for permits, $s_i^*(r)$, depends on its attitude towards risk and on the characteristics of the enforcement policy it faces. Hence, the equilibrium permit price will also depend on these factors.

As is true for pollutant discharge, the relationship between a noncompliant firm's permit demand and that of an otherwise identical compliant firm depends on the characteristics of the firm's subjective audit probability function. As established above, if ϕ_i is non-negative, $w_i^* \leq \bar{w}_i^*$. Since $w_i^* > s_i^*$ for a noncompliant firm, it follows that $s_i^* < \bar{w}_i^*$ when $\phi_i \geq 0$, that is, the permit demand of a noncompliant firm is lower than that of an otherwise identical compliant firm. (A compliant firm's permit demand is identical to its pollutant discharge level.) However, if ϕ_i is negative, $w_i^* > \bar{w}_i^*$. Although it is still true that $w_i^* > s_i^*$ for a noncompliant firm, this inequality no longer implies an unambiguous relationship between s_i^* and $\bar{w}_i^*$. Thus, it is possible when $\phi_i < 0$ that the number of permits demanded by a noncompliant firm is higher than that demanded by an otherwise identical compliant firm over some range of permit prices (i.e., $w_i^* > s_i^* > \bar{w}_i^*$).

The absence of an unambiguous relationship between the permit demands of noncompliant and compliant firms implies that no determinate relationship can be established between the equilibrium permit price in a market with noncompliant firms (r^*) and the equilibrium price in a market with otherwise identical compliant firms ($\bar{r}^*$). In particular, one cannot rule out the perverse possibility that the equilibrium permit price in a noncompliant market is higher than that in a compliant market. However, for the special case where ϕ_i is non-negative for all noncompliant firms in the TDP market, $r^* < \bar{r}^*$. This is shown below.

The equilibrium permit price in a compliant market is implicitly given by the equation $\sum_i \bar{w}^*(r) = s_T$, whereas the equilibrium price in a noncompliant market is given by $\sum_i s_i^*(r) = s_T$. When ϕ_i is non-negative, $\bar{w}_i^* > s_i^*$, hence

$$\sum_i \bar{w}_i^*(r^*) > \sum_i s_i^*(r^*) = s_T.$$

Since $\bar{w}_i^*(r)$ is decreasing in r (see (6) and recall $B_i'' < 0$), it follows that $r^* < \bar{r}^*$. Thus, when $\phi_i \leq 0$ for all noncompliant firms (a condition which is automatically satisfied when audit probabilities are constant or a function of violations size), noncompliance results in a lower equilibrium permit price.

The above analysis indicates that the principal effect of noncompliance in a TDP market is to alter the equilibrium permit price. An important implication of this result is that noncompliance on the part of even one firm in the market will affect the pollutant discharge levels of all the other firms through its impact on the equilibrium permit price.

Some of the results derived thus far are illustrated in Fig. 1 for a TDP market consisting of two risk-neutral firms with audit probabilities that, for ease of exposition, are assumed to be constant for all values of w_i and s_i. Each firm's marginal profit curve (B_i') is depicted along with its marginal expected penalty curve ($p_i F_i'$). Firm 1's pollutant discharge and permit holdings are measured from the left side of the box, and Firm 2's from the right side. The distance between the sides of the box represents the total number of permits issued, s_T.

If we ignore the possibility of noncompliance, the equilibrium permit price, $\bar{r}^*$, and the equilibrium allocation of permits, A_c, are determined by the intersection of the marginal profit curves. Since the firms are compliant, A_c also gives their pollutant discharge levels.

If we allow for noncompliance and consider the marginal expected penalty curves faced by each firm, it is clear that Firm 2 will be compliant in equilibrium since its marginal expected penalty evaluated at $v = 0$ is larger than $\bar{r}^*$, which is an upper bound on the equilibrium permit price in the presence of noncompliance. Firm 1, on the other hand, will choose to be noncompliant.

Trading in the market amounts to sliding the "branch" formed by the marginal expected penalty curves parallel to the horizontal axis. At the equilibrium depicted, Firm 1's marginal expected penalty is equated to its marginal profit, which, in turn, is equated to Firm 2's marginal profit. The common value of these three quantities is the equilibrium permit price, r^*. Clearly, $r^* < \bar{r}^*$ given the assumption that the audit probabilities are constant.

The equilibrium allocation of permits in the presence of noncompliance is given by A_n. Since Firm 2 is compliant, its pollutant discharge is equal to its permit holdings. However, Firm 1's discharge is given by $0_1 w_1^*$, which exceeds its permit holdings by the amount v_1. Both firms discharge more than they do in the absence of noncompliance.

Desirability of Equating Marginal Profits

In principle, when firms are perfectly compliant, TDP markets minimize the cost of achieving a given total discharge level since they ensure that marginal profits with respect to pollutant discharge are equated across firms. As we have seen, when firms are noncompliant, marginal profits may not be equated across firms. Only when the audit probabilities satisfy the condition in (5) over the relevant range do firms equate their marginal profits to the common permit price r. To conclude from this result that TDP markets may not be efficient when firms are noncompliant, it is necessary to establish that equating marginal profits is desirable even when firms are noncompliant.

In the perfect compliance case, the desirability of equating marginal profits is demonstrated by maximizing a social net benefit function of the form $[\Sigma B_i(w_i) - D(\Sigma w_i)]$, where D(.) captures the damages from pollutant discharges. The central authority is assumed to be able to directly control each firm's pollutant discharge. The first-order conditions for this problem establish that discharge levels are chosen such that marginal profits are equated across firms and set equal to the marginal damage from pollution.

The relevant benchmark maximization problem in the noncompliance case is considerably more complicated. It is now inappropriate to assume that the central authority can directly control each firm's pollutant discharge. However, it is reasonable to assume that the authority can indirectly control discharges by issuing *non*-marketable permits to firms: in effect, setting a discharge standard for each firm. Further control of discharges is provided by the authority's choice of enforcement policy. The objective function of the problem changes to the extent that it must also include the costs of enforcement.[4]

Characterizing the solution to the benchmark problem described above requires an analysis of the optimal enforcement policy, which is outside of the scope of this paper. The approach adopted here is to examine, instead, a sub-problem of the complete welfare maximization problem: one in which the enforcement policy is taken as given, and only the allocation of non-marketable permits is variable. The first-order conditions for the solution to this problem are a subset of those for the complete problem.

Since the enforcement policy is taken as given, enforcement costs are a constant and, for our purposes, can be omitted from the objective function. Hence, the objective function is identical in appearance to that for the perfect compliance case:

$$\max_{s_i} \sum_{i=1}^{n} B_i(\hat{w}_i^*) - D\left(\sum_{i=1}^{n} \hat{w}_i^*\right), \tag{7}$$

where $\hat{w}_i^* = \hat{w}_i^*(s_i)$ denotes a firm's optimal pollutant discharge level given a fixed number of permits, s_i. The first-order conditions for an interior solution to this problem are:

$$(B_i' - D')\frac{\partial \hat{w}_i^*}{\partial s_i} = 0 \qquad i = 1,\ldots,n. \tag{8}$$

These differ from the corresponding conditions for the perfect compliance case by the presence of $\partial \hat{w}_i^*/\partial s_i$. They imply, nonetheless, that it is desireable to equate marginal profits even when firms are noncompliant. We can therefore conclude that when firms are noncompliant, TDP markets are efficient only if $\phi_i = 0$ over the relevant range for all firms.

4. An Alternative Interpretation -- Effluent Taxes

A straightforward reinterpretation of the TDP model developed above allows us to examine the consequences of noncompliance for another pollution control policy widely advocated by economists--effluent taxes. To convert (1) to a model of a noncompliant firm facing an effluent tax, set s_i^o equal to zero, and redefine s_i to be the reported discharge on which taxes are paid and r to be the unit effluent tax. Noncompliance now takes the form of tax evasion, that is, a firm under-reports its discharges. If the tax evasion is not detected, the firm pays taxes only on its reported discharge s_i. However, if the firm is caught, it also pays a penalty of

$F_i(v_i)$ based on the discrepancy between actual and reported discharge $v_i = w_i - s_i$.

The apparent identity between the model of a noncompliant firm in a TDP market and that of a tax-evading firm must be interpreted with caution. In the effluent tax interpretation of (1), the effluent tax, r, is presumably exogenous and not a function of firm behavior.[5] In contrast, in the TDP market interpretation, r is the equilibrium permit price which, as we have seen, is influenced by noncompliance and is endogenous. However, if we consider the firm in a TDP market in a partial equilibrium setting, there is no substantive difference between the TDP market and effluent tax interpretations. Hence the results derived for the TDP model when r is fixed are those that would obtain for a firm facing an effluent tax. In particular, we can state that a tax-evading firm sets its marginal profits equal to the effluent tax only if its audit probability function satisfies the condition $\phi_i = 0$ over the relevant range. If this condition is met, noncompliance does not affect the quantity of pollutant discharged by a firm; it only implies lower tax revenues.

The above results are a generalization of those obtained by Harford (1978) and Storey and McCabe (1980) using simpler models of the tax-evading firm. Beavis and Walker (1983) incorrectly conclude from Harford's tax model that even in a TDP market a firm's pollutant discharge is unaffected by changes in the enforcement policy when discharge is deterministic (as is assumed here). Their error stems from overlooking the indirect link between a firm's discharge decision and the enforcement policy via the equilibrium permit price.

5. Implementing a TDP Market when Firms are Noncompliant

Detecting Violations

The results obtained above suggest that a system of transferable permits may not be as robust to noncompliance as an effluent tax. As we have seen, under certain conditions, the only effect of noncompliance on an effluent tax policy is to lower tax revenues; pollutant discharges are unaffected. This is not true for a TDP market given the same set of conditions. Implicit in this comparison of the two policy instruments, however, is the assumption that the ease with which violations can be detected is similar for an effluent tax and a TDP market. As is argued below, it is likely that, in practice, detecting violations would be easier with a TDP market.

Given the technical difficulties in obtaining accurate, continuous records of pollutant discharge, enforcement agencies generally must rely on intermittent estimates of a firm's pollutant discharge.[6] Typically, these estimates are based on data collected during on-site audits. Hence, firms usually are aware of when their discharges are being monitored. This does not necessarily pose a problem in the case of a TDP market since the amount a firm should discharge is known beforehand by the agency from the firm's permit holdings. As with an effluent standard, any observed discharge over the allowed amount signals a violation. If firms receive advance notice of audits, they may be able to buy additional permits for the duration of the visit, making it difficult for the agency to detect violations. But the scope for such strategies would be limited if audits are unannounced (as is assumed in the model) or permits can only be traded at certain times. David et al. (1980) have suggested quarterly auctions as a means of precluding such behavior. Thus, in a carefully designed TDP market, detecting violations should be no more difficult than under the existing policy of effluent standards.

Consider now the problem of detecting a violation when an effluent tax is employed. With a tax, the amount of pollutant a firm should discharge is determined when the firm reports (or formally records) its discharge, which occurs after the pollutant is discharged. As such, it is possible for tax-evading firms to adopt a strategy of accurately reporting discharges only when they are being monitored and thereby avoid being caught in *flagrante delicto*. Of course, there are limits to the extent to which this strategy would work. For instance, it is unlikely that firms could get away reporting near zero discharges when not monitored and levels substantially larger than zero when monitored. Nonetheless, the feasibility of this type of strategy implies that detecting violations is likely to be easier with a TDP market than with an effluent tax, given the existing technology for measuring pollutant discharges.

Containing the Effects of Noncompliance

A problem that is likely to be troublesome when implementing a TDP market is containing the effects of noncompliance. For an effluent tax or an effluent standard, changes in the compliance status of a particular firm do not necessarily affect the behavior of other firms: the effects of noncompliance are essentially localized. If a noncompliant firm revises its subjective audit probability as a result of, say, being caught and fined, this will not, per se, affect the decisions of other firms. Those that are compliant presumably will continue to be compliant, and those that are not may or may not alter their behavior. Of course, if the event is perceived as part of a larger crack down on noncompliance, the other noncompliant firms may revise their audit probabilities and alter their behavior accordingly.

In contrast, in a TDP market, the effect of noncompliance is transmitted throughout the market as a result of its influence on the equilibrium permit price. Changes in a single noncompliant firm's audit probability will, in general, alter the firm's demand for permits and, consequently, the equilibrium permit price. For a new equilibrium to be achieved, it may well be necessary for all firms, both noncompliant and compliant, to trade permits and adjust their discharge levels. This is especially true in markets with small numbers of dischargers, which are likely to be the rule rather than the exception.

Thus, given noncompliance, it may be difficult for a TDP market to achieve an equilibrium or maintain one, and firms may incur significant adjustment costs in modifying their discharge levels in response to changing permit prices. The significance of this problem will depend on the degree to which permit prices are affected by perceived changes in enforcement policy. In practice, permit prices may be far more sticky than models such as the one developed here lead us to expect.

Designing Trading Mechanisms to Accommodate Noncompliance

Undoubtedly, the stickiness of permit prices will depend on the mechanism adopted for effecting permit trades. It has been implicitly assumed above that the TDP market takes the form of a standard unregulated market where trades occur whenever two or more parties find them advantageous. However, a number of policy analysts have argued that to mitigate problems of market thinness, TDP markets should take the form of periodic auctions conducted by a central authority (e.g., Noll, 1982; David, *et al.*, 1980). A question that naturally arises is how noncompliance would be handled in such a market. Would a special auction be schedule whenever one or more firms are found in noncompliance to allow them to satisfy their possibly revised permit demands, or would they be required to wait until the next auction? In the former case, the auction-based market would function much like a conventional unregulated one. However, in the latter case, firms would effectively face a system of effluent standards between auctions, with each firm's standard given by its existing permit holdings. Hence, any efficiency properties of the TDP market, would be undermined between auctions. The significance of this problem clearly will depend on the frequency with which auctions are held, and the extent to which firms modify their discharge levels in response to perceived changes in the enforcement policy. It is likely that the problem can be mitigated in a carefully designed auction-based market.

6. Conclusions

Starting with a model of a noncompliant firm, I have examined the consequences of noncompliance for a system of transferable discharge permits. The analysis reveals that when firms are noncompliant TDP markets retain their efficiency property only under some fairly stringent conditions. The principal effect of noncompliance on the market itself is to alter the equilibrium permit price. Although it is likely that the equilibrium permit price is lower given noncompliance, it is difficult to rule out the possibility of a higher equilibrium price.

Via its effect on the equilibrium permit price, noncompliance on the part of any one firm, or group of firms, influences the behavior of all other firms in the market. This raises some important questions regarding the ability of a TDP market to achieve and remain in equilibrium when firms are noncompliant. It also implies that careful consideration must be given to noncompliance when designing the mechanism for effecting permit trades.

REFERENCES

Beavis, B. and Walker, M. (1983), "Random Wastes, Imperfect Monitoring, and Environmental Quality Standards," *Journal of Public Economics*, 21:377-387.

David, M., Eheart, W., Joeres, E. and David, E. (1980), "Marketable Permits for the Control of Phosphorous Pollution into Lake Michigan," *Water Resources Research*, 16:263-270.

Downing, P.B. and Kimball, J.H. (1982), "Enforcing Pollution Control Laws in the U.S.," *Policy Studies Journal*, 11:55-65.

Drayton, W. (1980), "Economic Law Enforcement," *Harvard Environmental Law Review*, 4:1-40.

Harford, J.D. (1978), "Firm Behavior under Imperfectly Enforceable Pollution Standards and Taxes," *Journal of Environmental Economics and Management*, 5:26-43.

Noll, R.G. (1982), "Implementing Marketable Emissions Permits," *American Economic Review, Papers and Proceedings*, 72:120-124.

Russell, C.S., Harrington, W. and Vaughan, W.J. (1986), *Enforcing Pollution Control Laws*, Resources for the Future, Washington, D.C.

Storey, D.J. and McCabe, P.J. (1980), "The Criminal Waste Discharger," *Scottish Journal of Political Economy*, 27:30-40.

[1] See Downing and Kimball (1982), Russell, et al. (1986), and Drayton (1980). The type of noncompliance being modeled here is what Drayton refers to as O&M noncompliance: day-to-day discharges in excess of the allowed amount.

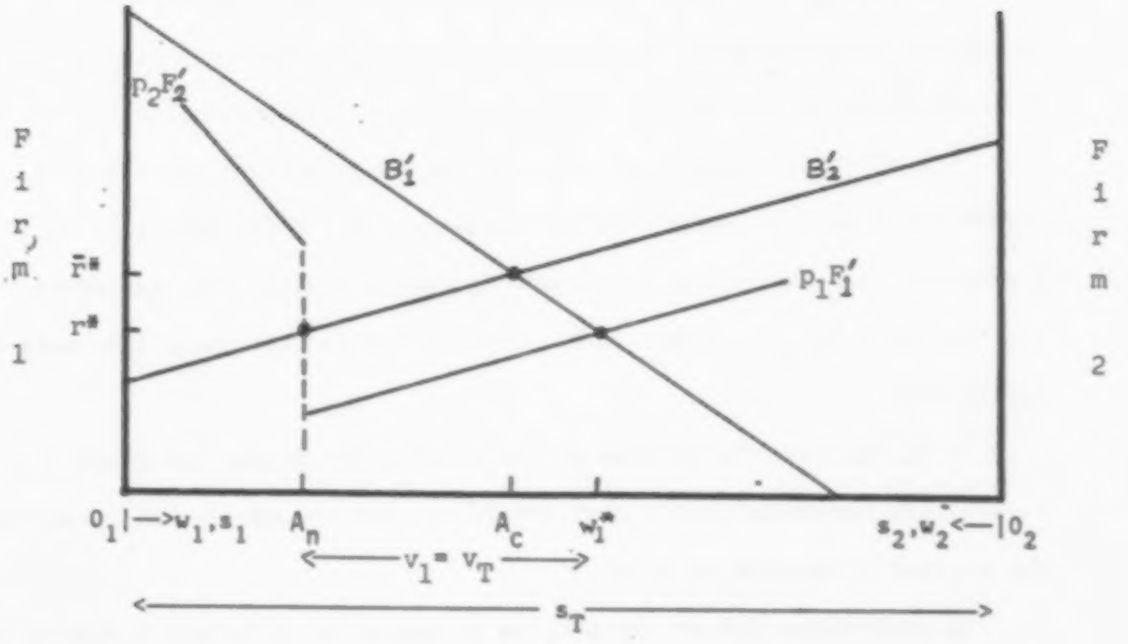
[2] Formally, $B^i(w_i) = \max_x[q_i x_i - c(x_i,w_i)]$, where x_i is the firm's product output and q_i is the price it receives for its output. The strictly convex function c(.) gives the firm's costs as a function of its output and pollutant discharge levels.

[3] The presence of w_i in the audit probability function does not contradict the assumption that levels of pollutant discharge are not known with certainty by the central authority prior to conducting audits.

[4] Formally, the additional decision variables are the audit policy parameters θ and the penalty policy parameters δ. Both these sets of parameters influence firms' pollutant discharge levels. The objective function is $\Sigma B_i(w_i) - D(\Sigma w_i) - C(\theta,\delta)$, where $C(\theta,\delta)$ represents the costs of enforcement.

[5] I am ignoring the problem of calculation the proper tax given uncertainty regarding firm's cost functions, and the associated incentives for strategic behavior by firms.

[6] A good discussion of the problem of measuring pollutant discharge is contained in Russell, *et al.*, (1986, pp. 76-86).



Firm 1
p_2F_2'
B_1'
B_2'
$\bar{r}^*$
r^*
p_1F_1'
Firm 2
$0_1 |\longrightarrow w_1, s_1$
A_n
A_c
w_1^*
$s_2, w_2 \longleftarrow| 0_2$
$v_1 = v_T$
s_T

Figure 1

TDP Market Equilibrium given Noncompliance

SESSION III

Empirical Models of Compliance and Performance of Enforcement Agencies

Jon G. Sutinen and John R. Gauvin, University of Rhode Island
"Enforcement and Compliance in the Massachusetts Lobster Fishery"

Wayne Gray, Clark University
Carol A. Jones, University of Michigan
"Longitudinal Analysis of the Impact of OSHA Health and Safety Regulations in Manufacturing"

Robert Rosenman, Washington State University
"Hitting the 'T': The Regulatory Loss of Soil Management Policy"

Wesley A. Magat, Duke University
W. Kip Viscusi, Northwestern University
"The Effectiveness of EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards"

PRELIMINARY DRAFT
DO NOT QUOTE

ENFORCEMENT AND COMPLIANCE IN THE MASSACHUSETTS LOBSTER FISHERY

by

Jon G. Sutinen
and
John R. Gauvin

Department of Resource Economics
University of Rhode Island

July 1987

Contribution No. __________ of the Rhode Island Agricultural Experiment Station.

Evaluation of enforcement programs ultimately depends on empirical measures of compliance and the factors which explain the incidence of compliance. Measuring compliance and its determinants is fraught with non-trivial problems. First, official data only partially measure non-compliance and usually cannot be used to estimate the overall incidence of compliance. Second, since compliance usually is a socially desirable trait, self-reports of compliance behavior are likely biased. And third, individuals' perceptions of certain conditions (e.g., the chance of being detected) may be more relevant than actual conditions when explaining compliance behavior.

In this paper we report on our recent attempt to overcome these and other problems. The subjects of our study are fishermen in the inshore commercial lobster fishery of Massachusetts. A survey of these fishermen was conducted during April and May, 1987, to collect basic data on enforcement and compliance in the fishery.

The purpose of this paper is two-fold: one, to present the methodology used for collecting the data and, two, to report our measures of compliance and of the determinants of compliance.

The paper is organized as follows. In the next section we describe the lobster fishery and the fisheries law enforcement program in Massachusetts. The second section explains the analytical framework and data collection methods used in the study. The third section presents some results of the survey, and the concluding section includes a discussion of the limitations of this work to date.

THE FISHERY AND ENFORCEMENT PROGRAM

Lobster is the most valuable inshore fishery in Massachusetts. In 1986, reported landings of lobster totalled nearly 15 million pounds valued at over $37 million. Approximately 88 percent of these lobsters were caught in inshore waters by 1400 fishermen with commercial inshore lobster licenses.

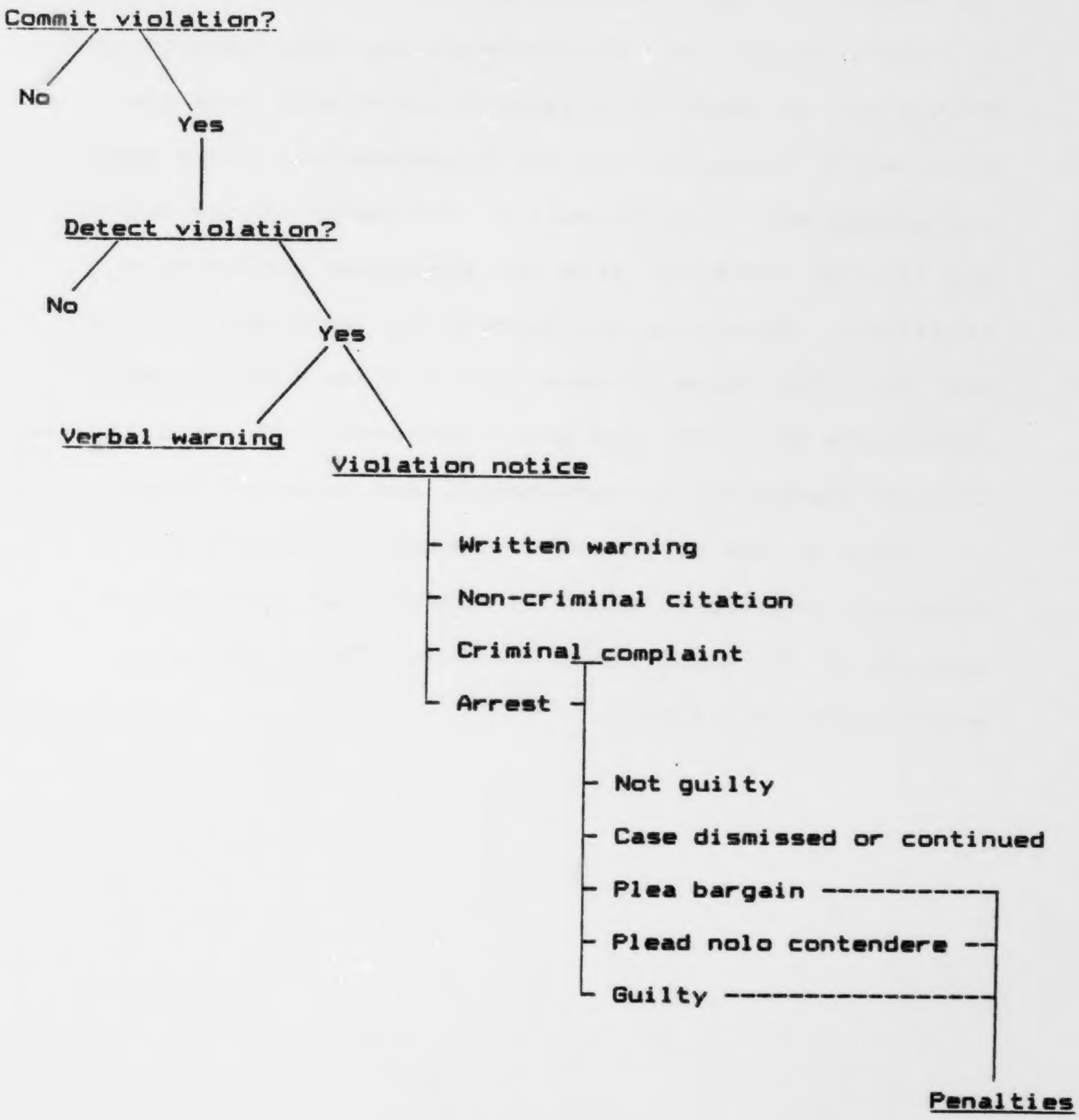
The fishery has been subject to a limited entry program since 1975. From 1975 to 1980 the number of inshore licenses was limited to about 1400, including hardship cases. In 1981, statutory changes allowed the number of licenses to increase. By 1986 there were approximately 1800 inshore commercial, 200 seasonal commercial, 700 offshore commercial, and 11,000 non-commercial licenses issued.

The Division of Environmental Law Enforcement is charged with enforcing Massachusetts fisheries laws. Among other things, these laws (i) set a minimum legal size for lobster (3 3/16 inches carapace length), (ii) prohibit the removal of eggs from lobsters and require egg-bearing lobsters to be immediately returned to the waters from which they were taken, (iii) prohibit anyone other than the owner from handling, destroying or molesting any lobster trap, and (iv) require a permit to possess and sell lobsters in the state. The Division enforces lobster and other marine fisheries regulations through its coastal enforcement bureau, consisting of six regions. In 1986 the coastal enforcement bureau had a staff of 42, 30 land-based wardens and 12 others operating two large patrol vessels. In addition to the patrol vessels, the Division owned ten small boats and contracted private aircraft to provide air patrols. The Division's complement of fisheries wardens is supplemented by deputized local police officers and harbor masters in most coastal towns.

The enforcement program employs three principal modes for detecting violations: air patrols, sea patrols and dockside inspections. Air patrols are used to monitor the location and general activities of the fishing fleet. Sea patrols are used to monitor the fleet and to board and inspect vessels on the water. Land-based wardens inspect landing and selling operations at the docks and conduct general investigations. When a warden detects a violation he either issues a written violation notice or a verbal warning. The written violation notice involves four possible outcomes: (1) a written warning; (2) a non-criminal citation, which carries a fine of no more than $100 payable in 21 days; (3) a criminal complaint; and (4) arrest with possible seizure of gear and vessel. The latter two outcomes, the criminal complaint and arrest, involve a hearing before a district criminal court of the State. The hearing process can result in five possible outcomes: (1) the fisherman may be found not guilty by the court; (2) the court may dismiss the case or continue it indefinitely; (3) the fisherman may plea bargain and settle out of court on a penalty negotiated with the Enforcement Division; (4) the fisherman may plead _nolo contendere_ and place himself at the mercy of the court; and (5) the court may find the fisherman guilty of a violation. Penalties are invoked for outcomes (3), (4) and (5). The range of penalties include imprisonment, forfeiture of vessel and gear, suspension of one's license to fish (temporary to permanent), and monetary fines.

This series of possible outcomes following a violation creates a "chain of deterrence" that is the essence of the enforcement program. The chain is illustrated in Figure 1.

FIGURE 1. THE CHAIN OF DETERRENCE


Commit violation?
No
Yes
Detect violation?
No
Yes
Verbal warning
Violation notice
Written warning
Non-criminal citation
Criminal complaint
Arrest
Not guilty
Case dismissed or continued
Plea bargain
Plead nolo contendere
Guilty
Penalties

METHODS

The basic model for evaluating a fisheries law enforcement program is developed in Sutinen (1986). In the model, compliance directly affects the size of the fish stock which, in turn, directly affects benefits from the resource, ceteris parabis. The incidence of compliance is directly related to the perceived probability of detection and conviction and the penalty for non-compliance, and inversely related to the expected gain from violating a regulation. The perceived probability of detection and conviction, in turn, is directly related to the resources and practices of the enforcement program. Penalties are assumed to be determined by a court in conjunction with the enforcement program. In this context, enforcement may be said to induce rather than produce compliance, since it affects the incentive structure (i.e., expected gains and losses) faced by individuals when deciding to comply or violate a regulation.

Data Collection

Primary data was collected to develop estimates of the variables for the compliance framework. The only secondary data available is the official data on detected violations maintained by the Division of Environmental Law Enforcement. The secondary data was rejected as a sole source because it is biased and incomplete. As explained in Sutinen and Hennessey (1986), these

> data merely measure the extent of detected noncompliance. That is, while the number of documented infractions is known, they are in fact only a subset of total noncompliance. If surveillance and monitoring were random, in some appropriate way, the levels of detected noncompliance could be extrapolated to the entire population to provide an estimate of the overall levels of noncompliance and compliance. Monitoring and surveillance--especially boardings and inspections--are not random, however, The typical enforcement program rationally focuses its surveillance and monitoring efforts on those vessels which most likely are not complying with regulations.... Other things equal, this concentration of surveillance and monitoring effort biases upwards the estimate of overall compliance. The bias cannot be adjusted for since the extent of concentration is not known and surely varies across fisheries and over time. (p. 195)

Since observation of compliance behavior would be extremely costly, if not impossible, we chose to survey a sample of the fishing population. (This approach was first applied by Blewitt, et al., 1987.) A mail questionnaire was the survey instrument used for data collection. The other survey instruments of personal and telephone interviews were rejected because they were expected to be overly time consuming or costly.

Another advantage is that data from mail questionnaires have been found to have less social desirability bias than data from personal and telephone interviews (Nederhof, 1985). Social desirability bias is introduced when subjects tend to deny socially undesirable behavior and to claim socially desirable behavior, and to report things which portray the subject favorably.

Clearly, in our survey, which inquires about a person's illegal behavior, the potential for social desirability bias affecting the data is great. To further minimize this bias we used the method of proxy subjects in the survey. That is, instead of asking a person about his/her compliance behavior, we asked respondents to report on another person's compliance behavior. Of course, the other person's anonymity was maintained. In his review of methods for coping with social desirability bias, Nederhof (1985) identifies the proxy subject's approach as yielding satisfactory data on behavior. The randomized response technique, currently a popular approach for coping with social desirability bias, was determined to be impractical for this survey.

The questionnaire was developed over a period of two months during which we consulted regularly with a panel of eight lobstermen. These meetings and several tests of the questionnaire were invaluable in designing our survey instrument. The final form of the questionnaire (see Appendix A) was designed to collect data on enforcement and compliance, and on gear losses, a major concern of most fishermen in the State.

The part of the questionnaire concerning enforcement and compliance was designed to provide the following information:

(1) respondents' personal experiences with and views of the enforcement program;

(2) estimates of the incidences of compliance in the fishery;

(3) descriptive profiles of typical compliers and violators;

(4) estimates of the monetary gains realized from violating regulations; and

(5) estimates of the probabilities of detection, prosecution and conviction as perceived by fishermen.

Sampling Procedures

The sample for the survey was drawn from the population of 1800 holders of licenses for the commercial, inshore lobster fishery in Massachusetts. As explained above, these fishermen account for 88 percent of the lobster landed in the state. Other participants in the fishery who were not surveyed include holders of commercial offshore lobster licenses, recreational trap license holders, and divers, both commercial and recreational.

The list of license holders, obtained from the Massachusetts Department of Fisheries, was sorted by towns, using postal codes. License holders in towns bordering each of our five study zones (see Fig. 2) were then grouped together. The study zones correspond roughly to the administrative regions used by the Division of Environmental Law Enforcement.

A sample of nearly 800 license holders was drawn for the mailing, distributed across zones as shown in Table 1.

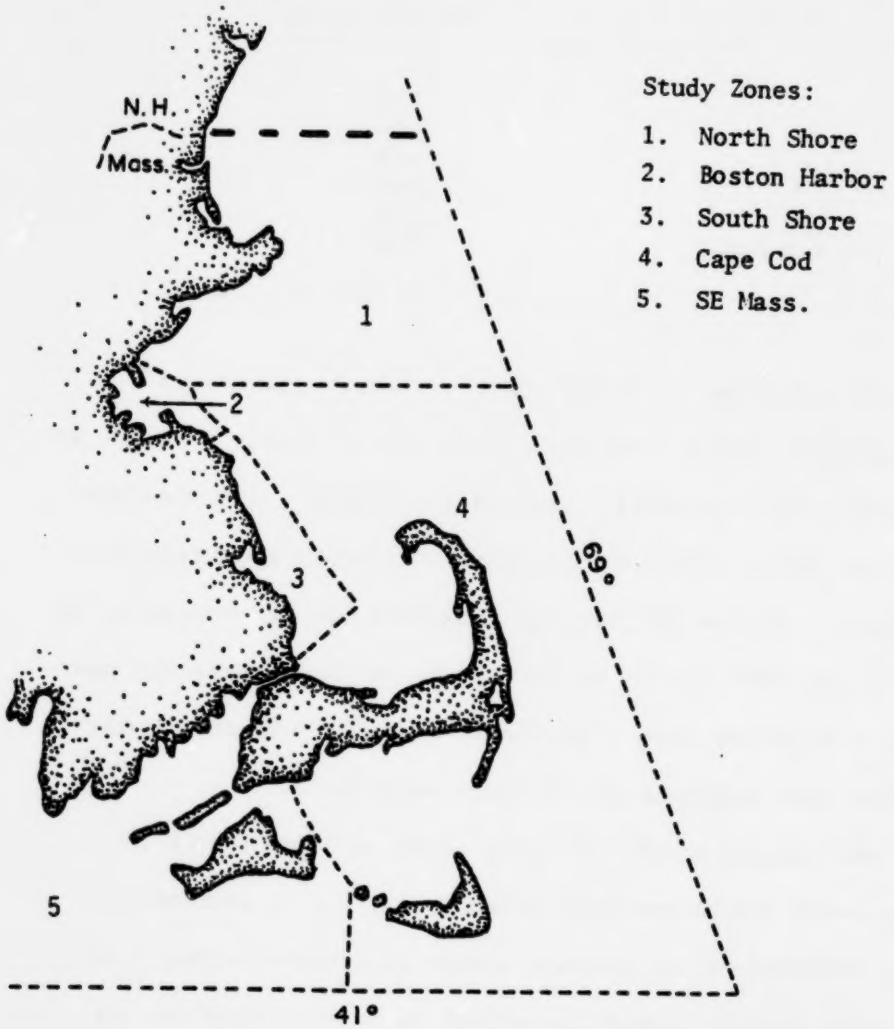

N.H.
Mass.
1
2
3
4
5
69°
41°
Study Zones:
1. North Shore
2. Boston Harbor
3. South Shore
4. Cape Cod
5. SE Mass.

FIGURE 2.
Location of zones for the study of enforcement in the Massachusetts inshore lobster fishery.

Table 1. Mailing Sample Distribution by Zone

Zone	Sample Size
1	166
2	166
3	166
4	165
5	135
all zones	798

A priori we expected a 25 percent response rate and selected sample sizes for each zone which would result in at least 40 usable questionnaires per zone. A principal objective of this study is to make reliable comparisons across zones. Since we had no information on the size of the population variance, we targeted 40 observations per zone merely because the t distribution is approximately normal when the degrees of freedom exceed 30.

The 800 questionnaires were sent out on April 3, 1987, and were followed two weeks later by a postcard reminding fishermen to return their questionnaires. A total of 272 usable questionnaires were returned to us within eight weeks, for gross response rate of 34 percent. Of these, 264, or 97% of the total returned, were received within four weeks of the mailing. (See Table 2).

Table 2. Cumulative Number of Questionnaires Received by Week

Week	Cumulative Number received	Percent of Total
1	82	30.1
2	127	47.1
3	230	84.6
4	264	97.1
5	268	98.5
6	269	98.5
7	270	99.3
8	272	100.0

Six different versions of the questionnaire were used. The sample of 800 was evenly divided between questionnaires inquiring about the behavior of fishermen who "usually comply with lobster regulations" and those who "frequently violate lobster regulations." These questionnaires are denoted as complier and violator questionnaires, respectively. The complier and violator questionnaires each has three different versions: one concerns taking and selling short (undersized) lobsters, a second concerns scrubbing and selling egg-bearing female lobsters, and a third concerns molesting other fishermen's gear. The distributions of the returned questionnaires by type and zone are shown in Table 3. The returned questionnaires are fairly evenly distributed across questionnaire types and geographic zones. The variation in response rates does not appear to be systematically related to questionnaire type or geographic zone.

Table 3. Distributions of Returned Questionnaires by Type and Principal Zone Fished

	Violator	Complier	Totals
Shorts	52	42	94
Egg-bearing females	42	43	85
Molesting	47	46	93
Totals	141	131	272

Zones	Mailed	Returned
	(%)	(%)
1	21	22
2	21	19
3	21	23
4	20	19
5	17	17
	100	100

RESULTS

The Respondents

Most respondents to our survey claim to be full-time lobstermen (73%), and to have worked in the Massachusetts inshore lobster fishery for more than five years (87%). In terms of where they fish, the respondents are distributed roughly proportionately across the five study zones.

There is near consensus among the respondents that enforcement is not adequate in the fishery (85%). Nearly two-thirds also claim that half or more of all enforcement is carried out by lobstermen alone, i.e. without the help or knowledge of enforcement officials.

Most respondents (88%) report seeing fisheries wardens one or more times working on the water or at the dock and a slight majority (52%) report being inspected at least once in 1986. When inspected, most respondents (79%) state their operations were never seriously disrupted by the inspection. Ten percent of those inspected report being issued violation notices during the year, most (64%) for short lobsters.

Respondents were asked to rate the State's enforcement program in five areas (see Table 4). The lowest rating concerns the Division's methods and use of equipment. Among other things, fishermen argue that the Division's two large patrol boats are too visible and too

Table 4. Ratings of Massachusetts' Law Enforcement Program (Relative frequencies of responses)

	Good (%)	Fair (%)	Poor (%)
Response rate to calls	34	41	25
Response time to calls	27	40	33
Follow-through on investigations	33	35	32
Dedication to effective enforcement	49	29	22
Effective methods and use of equipment	32	31	37
Fairness and neutrality	56	32	12

slow to be effective. The highest ratings are given for the Division's "dedication," and "fairness and neutrality." These ratings reflect an apparent respectful relationship between fishermen and wardens.

In summary, the respondents to our survey are deeply involved in the fishery, have substantial first-hand experience with the enforcement program, and express some serious concerns about the adequacy and effectiveness of enforcement in the fishery.

Compliers and Violators

In the process of developing and designing the survey our advisory group of lobstermen indicated that individual fishermen either know or have firm beliefs about which lobstermen usually comply with and which lobstermen regularly violate lobster regulations. Therefore, we decided to divide the population of lobstermen into two groups: those who usually comply, and those who regularly violate. Half of our questionnaires asked each respondent to describe an anonymous complier, and half asked for a description of an anonymous violator. The questionnaires were designed to obtain information on each group regarding their personal characteristics (but not enough to provide identification).

Our objective here was to construct and compare profiles of compliers and violators. This line of inquiry was, and continues to be, largely exploratory. The results regarding five personal characteristics are shown in Table 5. The numbers are the relative frequencies of responses. The selection of these five characteristics was based on the following conjectures:

(i) Fishermen who are less income-dependent on the lobster fishery have a weaker conservation motive and are more likely to violate management regulations.

(ii) Lobstermen with fewer years in the fishery are less likely to perceive or appreciate the long-term benefits of management and conservation measures.

(iii) Younger lobstermen face greater financial pressures (e.g., large mortgage payments) and are more likely to violate regulations for short-run gains.

(iv) Lobstermen with relatives in the fishery have a greater interest in the long-term health of the fishery and are less likely to violate management regulations

(v) Frequent violators are often in the fishery for short-term gains only.

Table 5. Profiles of Compliers and Violators

	Compliers	Violators
	(%)	(%)
Income		
<1/4	5	6
1/4 to 1/2	8	16
1/2 to 3/4	21	20
3/4 to all	66	58
Years fished		
<5 yrs	3	23
1-10 yrs	33	41
11-20 yrs	40	24
>20 yrs	24	12
Age		
<30 yrs	5	13
30-50 yrs	72	74
>50 yrs	23	12
Relative		
Yes	44	51
No	56	49
Short-term	6	30
Long-term	94	70

The survey results indicate that violators are perceived as being only slightly less income-dependent on the fishery than are compliers. Violators are reported to have been in the fishery for less time and to be slightly younger than compliers. There is little difference in family ties to the fishery between compliers and violators, and a few more violators than compliers are perceived to be in fishery for the short-term only. While generally consistent with our original conjectures, this evidence does not show strong or sharp differences between compliers and violators on these five characteristics.

Incidences of Violations and Illegal Gains

We estimate that 9 percent of all fishing trips involve landing short lobsters, 8 percent involve scrubbed egg-bearing lobsters, and 9 percent involve molesting other lobstermen's gear. These violations account for estimated annual landings of 614,000 pounds of short lobsters, 618,000 pounds of scrubbed egg-bearing lobsters, and 770,000 pounds from molesting others' gear. The value of these illegal landings is approximately $5.0 million, or 14 percent of reported landings in 1986. An average violator realizes about $13,500 in illegal income per year and complier about $2,400 (see Table 6).

Our estimates of violation rates and illegal gains are based on the survey data in Table 7 and a few strong assumptions. We assume, for example, that the violators and compliers described in the questionnaires are representative of their respective sub-populations. Since we did not collect data on compliers' illegal gains (for a trip with a violation), we assume that when compliers violate a regulation they realize the same gain per trip as that reported for violators. The data on gains per trip is for a typical trip in the peak fishing season and we assume all reported violations take place during the peak season.

Table 6. Annual Gains from Violations

	Violators	Compliers
Annual monetary gains per lobsterman from:		
Shorts	$ 4,536	$ 672
Egg-bearing	3,784	909
Molesting	5,151	1,414
Total	$13,471	$2,355
Estimated number of active lobstermen in each category	154	1,244
Annual monetary gains, 1398 active lobstermen, from:		
Shorts	$699,000	$ 836,000
Egg-bearing	583,000	963,000
Molesting	793,000	1,131,000
Total Estimated Gains $5,005,000		

Table 7. Violations Data for Violators and Compliers

	Violators	Compliers
Proportions of commercial lobstermen	.11	.89
Number of trips with violations of:		
Shorts	54	8
Egg-bearing	44	9
Molesting	51	9
Monetary gains per trip from:		
Shorts	$84	$84
Egg-bearing	$86	$101
Molesting	$101	$107

Probabilities

In the previous section we examined the extent of violations and the expected gains from those violations. Together with expected gains, the expected losses associated with violations determine the economic incentive to comply. The expected loss for a given violation is the product of the probability of detection and conviction and the penalty. The probability of detection and conviction is a function of the conditional probabilities for the series of outcomes in the chain of deterrence (Figure 1).

The probabilities include the following:

P_1 = P_r(detection on the water | a violation),

P_2 = P_r(detection at the dock | a violation),

P_3 = P_r(written violation notice | detection),

P_4 = P_r(criminal complaint | violation notice),

P_5 = P_r(no conviction | criminal complaint).

The overall probability of detection and conviction is given by

$$P_r\ (\text{Detection and Conviction}) = [P_1 + (1-P_1)P_2]P_3P_4\ (1-P_5).$$

In addition to affecting the compliance incentive, this series of probabilities makes explicit how each element or phase of the enforcement program is linked to the others. Given these probabilities, we can then consider the full tradeoffs of strengthening the various elements when evaluating the system.

To obtain estimates of these probabilities, we described in the questionnaires a series of hypothetical situations involving three separate types of violations (taking short lobsters, scrubbed egg-bearing lobsters, and molesting another fisherman's gear). The respondent was asked to estimate the odds of a specific outcome if the person described were in the hypothetical situation. An example of one of these questions is:

> Suppose this person, who you believe usually complies, for some reason has some SHORTS on board at the dock while he is unloading (wherever he unloads). What do you estimate the odds are that fisheries wardens would catch him with the shorts at that time? (Given enforcement as it was in 1986.)

The respondents were given a list of odds, ranging from one-in-one to one-in-1000, and asked to circle the odds closest to their answer.

The arithmetic mean is not a good measure of central tendency with this data because of its sensitivity to extreme values. The geometric mean, perhaps the best measure, cannot be computed until our software is modified. The mode would serve our purposes well were it not for several of the distributions using multi-modal. While not perfect, the median tends to give a fairly good indication of patterns in the data on probabilities. Medians of reported probabilities are given in Tables 8 and 9.

Table 8. Median Probabilities by Subject and Violation Types

Probabilities	All	Subject Type		Violation Type		
	naires	Violators	Compliers	Shorts	Egg-bearing	Molesting
Detection on the water, P_1 at the dock, P_2	 .002 .002	 .002 .01	 .002 .002	 .01 .01	 .001 .002	 .001 --
Violation Notice, P_3	.50	.50	.50	.50	.50	.50
Criminal Complaint, P_4	.20	.20	.10	.20	.20	.02
No Conviction, P_5	.50	.50	.20	.20	.50	.50
Detection and Conviction	.0002	.0006	.0002	.0016	.0001	.000005

Table 9. Median Probabilities by Zones

Probabilities			Zones		
	1	2	3	4	5
Detection					
on the water, P_1	.001	.002	.001	.002	.002
at the dock, P_2	.002	.002	.002	.01	.01
Violation Notice, P_3	.50	.50	.50	.50	.50
Criminal Complaint, P_4	.10	.20	.20	.20	.10
No Conviction, P_5	.20	.20	.20	.20	.50
Detection and Conviction	.0001	.0003	.0002	.001	.0003

As expected, the data reveal perceived probabilities of detection that are relatively small (around .002 in most cases). The perceived probability of receiving a written violation notice, given detection of a violation, is smaller than expected (.50). The Enforcement Division has a policy to issue a written notice, even for warnings, when any violation is detected. This modest value of P_3 may reflect wardens' lack of compliance with the written notice policy.

The odds of going to court if issued a violation notice are small (around .20 and less). This is surprising since the violations we described in our hypothetical situations are regarded as "serious" violations deserving a criminal complaint. The unusually small value of P_4 = .02 for molesting may indicate wardens in the field view the violation as less serious than the other types of violations.

The chance of escaping conviction are perceived as quite good in most circumstances (around .50). For reasons not yet clear to us, lobstermen who usually comply and those cited for taking shorts are perceived as having a higher chance of being convicted in court.

The breakdown of probabilities by zones reveals a few differences, but we have not yet had time to test for the statistical significance of these differences. Both the staff of the Enforcement Division and our advisory group of fishermen provided anecdotal evidence that differences exist in coverage and effectiveness among regions. If we can establish regions where below average weaknesses exist, the Enforcement Division is prepared to reallocate its resources or take other action to eliminate such weaknesses.

The overall probability of detection and conviction (calculated using median values) is generally in the neighborhood of .0002, equivalent to odds of one-in-5000. For violation of shorts the odds are greater at one-in-625; and smaller for molesting at one-in-200,000. Zone 4 has modestly better odds than other zones of detection and conviction at one-in-1000.

Similar calculations using means and modes result in larger overall probabilities of detection and conviction. Using arithmetic means gives a value of approximately .002, and using modes gives a value of approximately .001. Further analysis is required to select the proper measure and calculate the overall odds. At this point we conclude that the overall probability of detection and conviction appears to be no greater than .002. This magnitude of the probability implies that penalties and other costs must exceed $42,500 in order to offset the typical gain of $85 and deter the violations of taking short or egg-bearing lobsters.

CONCLUDING REMARKS

A framework for the empirical analysis of enforcement and compliance has been presented in this paper together with some descriptive statistics from a recent survey. While the methodology and resulting data seem promising, the current version of this paper has some significant limitations. First, we have not yet had sufficient time to conduct a complete statistical analysis of the survey data. More work needs to be done to measure the significance of our various statistics, and we still must estimate the behavioral relationships on compliance for which the survey was designed. Second, the issue of non-response bias has not yet been dealt with. We plan a random telephone survey of our original mailing list to identify potential differences between respondents and non-respondents. Third, we need to compare our results on fishermen's perceptions of probabilities with the actual outcomes recorded in the Division's records. These three sets of limitations will be addressed in subsequent revisions of this paper.

REFERENCES

Blewitt, E., W. Furlong, and P. Toews. 1987. Canada's Experience in Measuring the Deterrent Effect of Fisheries Law Enforcement. In, J. Sutinen and T. Hennessey (eds.), Fisheries Law Enforcement: Programs, Problems and Evaluation, Marine Technical Report 93, University of Rhode Island.

Fox, J.A. and P.E. Tracy. 1986. Randomized Response: A Method for Sensitive Surveys. Beverly Hills, CA: Sage Publications.

Nederhof, A.J. 1985. Methods of Coping with Social Desirability Bias: A Review. European Journal of Social Psychology 15:263-280.

Sutinen, J.G. and T.M. Hennessey. 1986. Enforcement: The Neglected Element in Fishery Management. In E. Miles, R. Pealy, and R. Stokes (eds.), Natural Resources Economics and Policy Applications. Seattle: University of Washington, Press.

Sutinen, J.G. 1986. Enforcement of the MFCMA: An Economist's Perspective. Paper presented at the annual meeting of the American Fisheries Society, Providence, Rhode Island.

Sutinen, J.G. and P. Andersen. 1985. The Economics of Fisheries Law Enforcement. Land Economics 61(4):387-397.

APPENDIX A

SECTION 5. Compliance Behavior

The information you provide in this section is very important for our analysis of improvements in the enforcement program. These are not easy questions. When you are asked to give your best guess, don't worry about being exact. The best guess of someone who knows the industry is often more accurate than one would expect.

First, think of all the commercial lobstermen you know who fish in the same zones that you fish.

24) Please estimate, to the best of your ability, what percent of the commercial lobstermen in your zone usually or always COMPLY with, and what percent usually or regularly VIOLATE the law and regulations pertaining to the lobster fishery. (PLEASE CHECK THE PERCENT GROUP THAT COMES CLOSEST TO WHAT YOU BELIEVE TO BE CORRECT)

Regularly VIOLATE	Usually COMPLY
____ zero percent	____ zero percent
____ 1 percent	____ 1 percent
____ 3 percent	____ 3 percent
____ 5 percent	____ 5 percent
____ 10 percent	____ 10 percent
____ 20 percent	____ 20 percent
____ 50 percent	____ 50 percent
____ 80 percent	____ 80 percent
____ 90 percent	____ 90 percent
____ 95 percent	____ 95 percent
____ 97 percent	____ 97 percent
____ 99 percent	____ 99 percent
____ 100 percent	____ 100 percent

Now, think of one commercial lobsterman in your zone who you believe frequently VIOLATES lobster regulations. DO NOT identify this person.

Please answer the following questions about this person. Please do not write this person's name anywhere on the questionnaire or we will hav [illegible] iscard your questionnaire. Keep the SAME person in mind for all of the [illegible] stions.

25) Roughly how old is this person? (CHECK ONE)

____ Under 30 years of age

____ 30 to 50 years

____ Over 50 years

26) What zone(s) did this person fish in 1986?

Zone(s) ________ (SEE MAP)

27) Roughly, how many days did this person fish for lobster in 1986?

____ (DAYS)

28) About how much of this person's income came from commercial lobster fishing in 1986? (CHECK ONE)

____ less than one-fourth

____ one-fourth to one-half

____ one-half to three-fourths

____ three-fourths to all

29) About how many years has this person fished in the Mass. inshore commercial lobster fishery? (CHECK ONE)

____ less than 5 years

____ 5 to 10 years

____ 11 to 20 years

____ more than 20 years

30) In your opinion, is this person fishing in the Mass. inshore lobster fishery for "the short term" or for "the long term"? (CHECK ONE)

____ Short Term

____ Long Term

31) Does this person have any relative(s) working in the Mass. lobster fishery?

____ Yes; ____ No; ____ Don't know.

32) To the best of your knowledge, has this person EVER done any of the following?

(1) Taken more than 2 shorts .... ___ Yes; ___ No; ___ Don't Know

(2) Taken or scrubbed egg-bearing females .... ___ Yes; ___ No; ___ Don't Know

(3) Molested someone else's gear .... ___ Yes; ___ No; ___ Don't Know

(4) Fished with an invalid license .... ___ Yes; ___ No; ___ Don't Know

33) Has this person ever been given a Violation Notice by fisheries wardens?

_____ Yes; _____ No

34) If you answered yes to any part of question 32, please estimate how many of this person's trips involved the following offenses in 1986.

(1) Taking more than 2 shorts .... ________ (NUMBER TRIPS)

(2) Taking or scrubbing egg-bearing females .... ________ (NUMBER TRIPS)

(3) Molesting someone else's gear .... ________ (NUMBER TRIPS)

(4) Fishing with an invalid license .... ________ (NUMBER TRIPS)

35) Please circle the month(s) when this person most likely committed each type of offense in 1986.

Shorts:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Eggers:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Molesting:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Invalid License:	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC

36) How much EXTRA money per trip to you think this person makes on a typical trip in the peak fishing season by

Selling shorts caught in his traps .... $________ extra per trip

Scrubbing and selling eggers caught in his traps .... $________ extra per trip

Molesting other's gear .... $________ extra per trip

37) Suppose this person you are thinking of has some SHORTS on board at the dock while he is unloading (wherever he unloads). What do you estimate the odds are that fisheries wardens would catch him with the shorts at that time? (Given enforcement as it was in 1986.)

(CIRCLE THE ANSWER YOU FEEL IS CLOSEST TO THE ODDS)

1 in 1	1 in 5	1 in 10	1 in 20
1 in 50	1 in 100	1 in 500	1 in 1000

38) Now think of a slightly different situation. This time the person is out fishing on the open water and has shorts in his tanks. What is your best guess of the odds that this person would be caught by fisheries wardens at this time with shorts? (Given enforcement as it was in 1986.)

(CIRCLE THE ANSWER THAT COMES CLOSEST TO THE ODDS)

1 in 1	1 in 5	1 in 10	1 in 20
1 in 50	1 in 100	1 in 500	1 in 1000

39) Suppose this person is caught by fisheries wardens with ten (10) obviously short lobsters banded and in his tanks. Given who he is (his reputation) and the way enforcement generally operates in your area, what do you think are the odds that he will be given a VIOLATION NOTICE instead of just a verbal warning?

(CIRCLE THE ANSWER THAT COMES CLOSEST TO THE ODDS)

1 in 1	1 in 2	1 in 5	1 in 10
1 in 20	1 in 50	1 in 100	1 in 200

A VIOLATION NOTICE can result in one of four outcomes:

(1) a Written Warning with no fine or penalty;

(2) a Non-Criminal citation which carries with it a fine of up to $100, payable within 21 days;

(3) a Criminal Complaint which involves a hearing in court, where a judge determines guilt or innocence; or

(4) Arrest of the violator, which may include seizure of gear, vessel, or motor vehicle.

40) Now think of the person you had in mind who frequently violates. Knowing who he is (his reputation and whether he has been caught by officials before), what is your best guess of the odds that he would be given a CRIMINAL COMPLAINT (instead of the other notices described above) for getting caught with the ten short lobsters?

(CIRCLE THE ANSWER THAT COMES CLOSEST TO THE ODDS)

1 in 1	1 in 2	1 in 5	1 in 10
1 in 20	1 in 50	1 in 100	1 in 500

If someone is given a CRIMINAL COMPLAINT that person must go to court and five possible outcomes can result:

(1) Guilty - the court hears the case and finds the defendant guilty of the charge or charges.

(2) Plead to Sufficient Facts (Nolo contendere) - guilt is admitted, defendant places himself at themercy of the court.

(3) Plea Bargain - case is settled out of court with a penalty agreeable to both the plaintif and the defendant.

(4) Not Guilty - the court finds the defendant innocent of the charge or charges.

(5) Case Dismissed or Continued - no finding is made by the court and no punishment results.

41) Suppose the person you have in mind is given a Criminal Complaint for the ten short lobsters. From your knowledge of how these matters have worked out in the past, what do you estimate the odds are that this person would get off (found Not Guilty or his case Dismissed or Continued)?

(CIRCLE THE ANSWER THAT COMES THE CLOSEST TO THE ODDS)

1 in 1	1 in 2	1 in 5	1 in 10
1 in 20	1 in 50	1 in 100	1 in 500

42) Suppose this person is found guilty of a Criminal Complaint for the ten short lobsters. What is your best guess of the odds that this person would receive a License Suspension of 60 or more days?

(CIRCLE THE ANSWER THAT COMES THE CLOSEST TO THE ODDS)

1 in 1	1 in 2	1 in 5	1 in 10
1 in 20	1 in 50	1 in 100	1 in 500

43) Suppose this person were seen by other commercial lobstermen taking ten obviously short lobsters. Given this person's reputation and the way enforcement works in your zone(s), what do you believe the odds are that other lobstermen would take direct action against him (instead of just warning him or informing fisheries wardens)?

(CIRCLE THE ANSWER THAT COMES THE CLOSEST TO THE ODDS)

1 in 1	1 in 5	1 in 10	1 in 20
1 in 50	1 in 100	1 in 500	1 in 1000

Now, please use the enclosed stamped envelope to mail this questionnaire to the University of Rhode Island, Department of Resource Economics, Kingston, RI 02881.

THANK YOU FOR YOUR HELP

LONGITUDINAL ANALYSIS OF THE IMPACT OF OSHA HEALTH AND SAFETY REGULATIONS IN MANUFACTURING

by

Wayne Gray,
Clark University

and

Carol A. Jones,
University of Michigan

DRAFT: not for citation without permission from the authors.

Acknowledgments: This paper was prepared for presentation at the AERE Workshop on Monitoring and Enforcement, August 1987. We would like to thank the Occupational Safety and Health Administration for providing us with the data to analyze in this paper. We also would like to acknowledge the useful comments of John Bound, John Scholz, and participants in the session on OSHA Enforcement during the Law and Society Annual meetings, June 1987, where we presented a previous version of this paper. We alone are responsible for any errors.

Longitudinal Patterns
# of Compliance with OSHA Health
and Safety Regulations in the Manufacturing Sector

1. Introduction

During the 1970s the United States experienced a dramatic expansion in public controls on private behavior designed to upgrade environmental, occupational and product safety. To improve occupational safety and health, the Occupational Safety and Health Administration (OSHA) was established in 1970. Numerous studies have subsequently examined the impact of OSHA on safety performance in the U.S. As McCaffrey et. al. [7] pointed out, however, the research results pose a puzzle. On the one hand, "intensive studies of individual firms suggest that OSHA does improve safety performance and does increase safety-related investment" [p. 198]. On the other hand, "regression-based studies find fairly consistently that factory inspections by OSHA do not reduce industrial injury rates or increase [safety-related] investments."[p. 198][1]

The discrepancies may occur because the effect of OSHA is relatively small and is swamped in the regression-based studies by statistical problems. Careful consideration of the econometric studies suggests that many are subject to specification and measurement problems which are likely to bias estimates of OSHA's impact. Independent analyses conducted by safety professionals suggest that only a _small percentage_ of accidents could be prevented by compliance with OSHA standards.[2] If, as a result, OSHA's effect on the _total_ accident rate is small, the estimation problems in the econometric studies could swamp the effect.

---

[1] For case studies, see cites [2], [3], and [6]; for regression-based studies, see [1], [7], [8], [11], [12], and [13].

OSHA health standards are widely believed to be more efficacious in reducing future occupational disease than safety standards are in reducing accidents. However, virtually all econometric studies to date have ignored OSHA's impact on health, because data on health effects are not readily available. This omission may produce substantial underestimates of OSHA's total effect on workplace quality.

In this paper, we analyze a unique plant-level longitudinal dataset that allows us to study OSHA's impact on health as well as safety. Derived from OSHA's enforcement MIS, the dataset includes information on citations for violating OSHA health and safety standards and on the levels of worker exposures to hazardous substances.

In this section, we briefly review previous studies of the impact of OSHA, to provide a context for our analysis. The regression-based literature generally can be categorized into two methodological types. In the first set, a specific deterrence model is estimated for a plant-level data set usually covering a one- or two-year period. In the second set, a general deterrence model is estimated for industry-level data which may span the full period of OSHA's activities. The standard measures of enforcement are different for the two models: the first method estimates the impact of the occurrence of an inspection; the latter method estimates the effect of expected penalties (probability of inspection times the fine).

---

[2] See Mendeloff [8, p. 86] for citations and discussion.

By focusing on the ex post reaction, the specific deterrence model emphasizes the role of inspections in providing information and, perhaps, a "management shock" stimulating the company to re-evaluate safety and health investments. The general deterrence model, on the other hand, emphasizes the importance of ex ante expectations of non-compliance penalties in promoting compliance with the regulations. These ex ante expectations may be implicit in the response to an inspection in a specific deterrence model, but the model does not allow us to measure directly the effects of policies to change expectations (for example, increasing the probability of inspection or the fine.) However, none of the current work explicitly incorporates expectations of future detection and sanctions, conditional upon past inspection-compliance performance. This limitation is common to both methodologies. However, analyzing a specific deterrence model with longitudinal plant-level data makes it possible to differentiate plant responses on the basis of past enforcement actions.

Despite their differences, the two models both define the effect of OSHA solely in terms of responses to enforcement actions. Case studies suggest that OSHA also engenders a "general awareness" or "existence" effect, in which the existence of OSHA raises concern about occupational safety and health within corporate management and among production workers. The awareness provides additional leverage for safety directors in gaining access to corporate resources and provides leverage for unions in contract bargaining. Only one study has tested the more general model of OSHA impact, the "existence" effect [8]. In this study, Mendeloff derived injury rate predictions for the post-OSHA period 1971-1975 from pre-OSHA relationships during 1948-1970. These predictions were then compared against observed injury rates. [We discuss the results of the study below.]

As noted above, various specification and measurement problems plagued the studies. In the industry-level studies, the effect of OSHA is likely to be underestimated due to the aggregation of injury and enforcement data across diverse plants within each industry. Most studies are estimated with data aggregated to the 4- or 2-digit SIC level, except for the specific deterrence studies which analyze plant-level data using very restrictive models of OSHA impact.

Mendeloff [8] demonstrated that the more important aggregation problem, however, is aggregation over different categories of injuries. Almost all studies, both on the plant and industry levels, have employed total accident rates. In contrast, Mendeloff analyzed the determinants of disaggregated categories of injuries. For categories independently identified by safety professionals as having a *high proportion of injuries caused by detectable violations of OSHA standards*, observed injury rates were approximately 20% lower than injury rate predictions. The analysis of disaggregated injury rates highlights the substantial effect OSHA *does* have on *selected* injury groups.[3] Because the injuries associated with violations represented a small share of total injuries (5-30% by different criteria), the effect on *total* injuries was small. Mendeloff's methodology is not directly comparable to the methodology of the other studies, as noted above. His atypical findings that *OSHA is effective in selective circumstances* may be due in part to modelling an "existence" effect for OSHA.

---

[3] Mendeloff faced the difficulty of distinguishing the "OSHA existence" effect on reducing accidents during the 1970s from the effects of other contemporaneous occurrences (not captured in the exogenous variables in the equation.) The most important omission in the model is workers' compensation benefits: his model does not capture the rapidly increasing workers compensation benefits during this time, which provides incentives to firms to reduce accidents but also provides incentives to injured workers to report more of their accidents. In empirical studies, the reporting effect generally swamps the deterrence effect on accident rates. If this pattern holds in this context, the omission yields an *underestimate* of the OSHA deterrence effect with his study methodology.

However, the use of disaggregated categories of accidents is probably the more important factor, because his results are comparable to those in previous studies for total injuries.

Bartel and Thomas [1] estimated a general deterrence industry-level model, in which they estimated separately the effect of OSHA enforcement on company compliance and the relationship between compliance and injury rates.[4] The new result in their analysis is that increasing enforcement intensity appeared to be positively associated with greater compliance. Because they used aggregate injury rate data, it is not surprising that the relationship between OSHA violations and the injury rate was small and imprecise in their study.

Many studies are subject to a variety of estimation problems in addition to aggregation bias. The exclusion of workers' compensation variables will tend to produce underestimates of the impact of OSHA in either general or specific deterrence models explaining accident rates[5] (though not with the Mendeloff methodology, as discussed in footnote 1.) Some studies have suggested that injury rates are measured with error.[6]

---

[4] The third equation in their model characterizes the determinants of inspection rates. We will not discuss that part of the analysis, because it is not germane to our concerns here.

[5] Russell [10] and Robertson and Keeve [9] demonstrate the importance of this mis-specification.

[6] McCaffrey et. al. [7] discuss this issue.

Furthermore, the data series on inspection rates and injury rates are highly auto-correlated, which poses a serious challenge in identifying causal effects, particularly in models with lagged enforcement variables. Such specification problems could produce unstable parameter estimates across time periods.[7] As noted above, these several estimation problems could prevent detection of the presumed small effect of OSHA on total accident rates.

The interesting question that has not been addressed in the literature is, what is the effect of OSHA on occupational health quality produced in firms? Health standards are widely hypothesized to be more closely associated with the production of health quality than the safety standards are with safety quality. The fundamental requirement of the health standards is to reduce exposures below the permissible exposure limit [PEL]. Given that exposure levels are generally considered to be a reasonable proxy for future incidence of occupational disease, the association between the standard and occupational disease prevention appears to be very close. By focusing solely on safety, analysts have ignored the more likely locus of OSHA impact. Furthermore, the Bartel and Thomas results showing that enforcement intensity is strongly associated with safety compliance should be a good portent of the potential effects of OSHA on health quality, given that OSHA health standards are closely associated with health quality.

---

[7] This problem may be the cause of Viscusi's [12] unstable estimates of the OSHA effect in his recent article.

The first analysis reported in this paper assesses the impact of OSHA enforcement activities in both the health and safety areas. We have created a unique dataset with longitudinal records for individual plants for 1972-83. With this dataset, we examine the inspection history of individual plants to determine whether the number of OSHA citations has declined with repeated inspections. Because no plant-level injury rate information [disaggregated or not] is available to us, we cannot attempt to replicate the Mendeloff results regarding the relationship between enforcement and injury rates. However, we do have a direct proxy for the future incidence of occupation disease: the measures of workplace exposures to health contaminants collected during health inspections. Furthermore, we have argued that compliance with the health standards is also a reasonable secondary proxy for health-related performance, unlike for safety-related performance.

The second analysis focuses solely on OSHA's impact on health. We examine the impact of OSHA enforcement on two measures of health performance: citations of OSHA standards and worker exposures to hazardous substances.

The next section of the paper presents a simple model of enforcement and compliance. The third section describes the data used in the analysis. The fourth section presents the results of the two analyses, and the final section discusses the results.

2. The Model

The major actors in our model of workplace safety and health are OSHA and private companies. OSHA sets standards, inspects plants, and issues citations and penalties when violations of the standards are detected. Each company is assumed to choose a level of compliance with the standards. The compliance level, in turn, has implications for the workplace safety and health quality level. In this paper we focus on company responses to agency enforcement (and the implications of the violation level for workplace "quality".)

Following the tradition in the plant-level analysis in the OSHA literature, we employ a specific-deterrence framework. The longitudinal structure of our data allows us to provide a far richer picture of longitudinal patterns of inspection and compliance through the first 12 years of the agency's history than possible in previous studies. To measure agency enforcement, we employ dummy variables indicating the sequence number of the inspection from the first through the fifth inspection [SEQNUMj, j=1,...,5] and a continuous variable denoting each additional inspection after the fifth [SEQNUMC].[8] At any given point in time, the number of previous inspections signals the intensity of (past) enforcement.[9] As noted earlier, the initial inspections may disseminate information about OSHA requirements and may provide a "management shock" to action. These functions are potentially crucial. Though the specific deterrence framework explicitly models the response of firms to past inspections, we implicitly assume that at least part of the response is motivated by anticipation of penalties in future enforcement actions. Though penalties for initial violations tend to be very low, the penalty schedules for repeat and willful violations cited in subsequent inspections are substantially higher.

---

[8] The variable equals total inspections - 5, for those plants with more than five inspections, and 0 otherwise.

We assume each company chooses a level of compliance with workplace health and safety standards, based on expected benefits and costs of the expenditures. The costs of compliance include the expenditures on machinery, protective clothing and equipment, and the foregone revenues associated with a slower workpace or alternative operating procedures necessary to comply with the standards. The benefits of expenditures to control hazards include avoidance of expected non-compliance penalties, as well as satisfaction of the firm's preferences for law-abiding behavior. We do not have direct measures of the benefits and costs; rather, we assume that they vary with the employment size

---

[9] We implicitly control for the elapsed time between inspections by including a variable measuring total number of inspections a plant receives during the panel period (NINSP). The role of this variable will be discussed further below.

(ESTSIZE) and industry (SIC) of the plants.

The OSHA enforcement data indicate the number of violations cited and the penalties levied in inspections. Because the citation data form the basis for our violation measures, we consider several factors which may affect the consistency of the relationship across inspections or through time between "true" violations and our measures of violations. The number of citations provides a measure of violations of all OSHA standards and the occurrence of worker exposure measurements provides an indicator of violations of OSHA permissible exposure limits for hazardous substances.[10]

First, different Administrations may vary in the rate at which enforcement officers choose to cite various types of violations. For example, the agency was widely criticized in its initial years for extensively citing trivial standards. OSHA substantially reoriented its policy in 1977, emphasizing detection of violations of more serious standards.

---

[10] See our detailed study of the exposure data [5], for further discussion of sampling and reporting issues associated with the exposure data. An in-depth study of records in two OSHA offices indicated that compliance samples were taken in 64% and 76% of all health inspections but samples were reported in approximately half of those inspections. However, in our study the lack of reporting appeared to be random: Contrary to our prediction, the distribution of severity levels for all samples taken (as indicated in area office files) was approximately the same as the distribution of severity levels for all samples reported in the MIS by area offices. We are unable to make any definitive inferences about the workplaces in which no samples were taken.

We control for variations in citation policy across Administrations with dummy variables for each year.[11]

Second, the origin of each inspection (complaint, follow-up, general schedule, accident) affects how much of an establishment is inspected, and therefore affects the likelihood that violations will be detected. General schedule inspections involve the broadest coverage of the workplace; complaint and follow-up inspections are generally narrowly focused on the subjects of the complaint or of past violations, respectively. To control for these variations in the relationship between "true" violations and citations, we will include dummy variables for inspection origin in the equation.

Third, the relationship between measured violations (citations) and "true" violations is affected both by the skill of the inspectors in detecting violations and by the skill of plant managers in practicing subterfuge. It is likely that both have increased through time. The net effect is indeterminable, but we do not expect that variations in relative detection/subterfuge skills across time periods will produce serious bias. Finally, the relationship is affected by the relative shares of detectable and non-detectable violations. Not all violations are detectable. Non-detectable violations include short-term stochastic events that are unlikely to be detected because inspections are relatively infrequent. If non-detectable violations are reduced by OSHA inspections less effectively than detectable violations, the observed decrease in detectable violations will overstate the decrease in "true" violations through time.

---

[11] However, the penalty policy associated with repeat and willful violations (which represents an important source of the deterrent threat associated with repeat inspections) changes across Administrations, so these variables are capturing a variety of effects. As a result, the coefficients must be interpreted with care (since they incorporate both agency citation behavior and deterrent effects on violations by firms.)

Note we only observe the violation level when an enforcement officer inspects an establishment. Forty percent of the plants are inspected only once, yet 2% [N=2667] have experienced 10 or more recorded inspections. This process of selecting observations into the sample potentially leads to serious sample selection bias. If OSHA is effective in targeting high violation plants for inspection, plants repeatedly selected for inspection are likely to be different from infrequently or never inspected plants. Enforcement officials may use information in their targeting process that is not available to the researcher, such as the local reputation of the plant. If plants that are inspected different numbers of times are consistently different from one another in ways we [the researchers] cannot measure, then estimates of the deterrent effect of inspections. Because we do not model directly the non-random selection process, they may be biased estimates of the effect of inspections in the universe of plants. In essence the problem occurs because we do not observe each plant in the sample the same number of times.

In our analysis, we will allow the estimates of company responsiveness to inspections to vary across the sub-samples [defined by the total number of inspections.] As discussed above, our estimates of the deterrent effects will be conditional upon the sample. If the deterrent effects are consistent across subsamples, however, we will have greater justification for making inferences beyond the sample.

We control for the selection effect by incorporating a variable measuring the total number of times a plant is inspected during the period, [NINSP], and by interacting the inspection [SEQNUM] dummy variables with NINSP. As with SEQNUM, for NINSP we employ dummy variables equal to 1 through 5, and then a continuous variable for additional inspections. For the first five inspections, we include SEQNUM and NINSP interactions. For inspections after the fifth, we assume the inspection-related deterrent effects are the same across the NINSP sub-samples. We also run a version of the regressions without the interaction terms.

All the other explanatory variables are also incorporated as dummy variables.

The estimating equation then becomes:

$$(1)\ V_{ij} = a_0 + \sum_{k}\sum_{q} a_{1,k}\mathrm{SEQNUM}_{ik}{*}\mathrm{NINSP}_{i,q} + a_2\mathrm{SEQNUMC}_{ij} + \sum_{m} a_{3,m}\mathrm{INORIGIN}_{ij,m} + \sum_{n} a_{4,n}\mathrm{INYR}_{ij,n} + \sum_{o} a_{4,o}\mathrm{SIC}_{j,o} + \sum_{p} a_{5,p}\mathrm{ESTSIZE}_{j,p} + \sum_{q} a_{6,q}\mathrm{NINSP}_{i,q} + a_7\mathrm{NINSPC}_{i} + e_{ij}$$

where the subscript i refers to plant i, subscript j refers to inspection sequence number j, and k, m, n, o, p, and q are indices of the dummy variable sequences. We predict that as the number of inspections of a plant increases, the numbers of citations in subsequent inspections declines. If unmeasured plant-level effects result in heteroskedastic errors, OLS estimators for equation (1) will be consistent but not efficient. Due to the extremely large sample size, however, inefficiency of the estimators is not a major issue.

We will estimate equation 1 both for the number of violations cited and, in the health analysis, for the number of worker overexposures.

3. Data

The source of data for the analysis is OSHA's enforcement Management Information System, used by the agency to track agency enforcement and company compliance performance. The version of the MIS data obtained for this study includes the 299,295 federal inspections performed in manufacturing establishments between 1972 and the middle of 1983.[12] The MIS provides information about OSHA's enforcement actions identifying which standards are cited and what penalties are levied. It also identifies the inspection origin (complaint, general schedule, accident, or followup), category (safety and health), and date of occurrence. Information is also provided about establishment characteristics, including the number of employees and industry, as well as individual plant identifiers.

---

[12] Not included in the data are those few inspections done in 1971 and 1972 before the MIS was operational and inspections performed in "state plan" states, in which state authorities have taken over responsibility for enforcement and OSHA only performs occasional inspections.

In order to create longitudinal records of plant inspection histories, Gray [1986] matched all inspections of individual establishments using establishment-level identifiers.[13] The weights given to agreement and disagreement on the data items used for matching determine the relative shares of Type I and Type II errors in linkage. Because of the variation in coding of establishment data over time (including errors in data entry), there are almost certainly cases in which inspections of the same establishment are not identified as such. It is also possible (though less likely given the structure of the weights) that inspections of different establishments are mis-identified as repeat inspections of a single establishment.

Table A1 describes the means and standard deviations of the first analysis sample: all health and safety inspections. The matching procedure identified 115,236 plants in the sample. Approximately 42% of the plants were inspected only once. The conditional probabilities of subsequent inspections were approximately 60%, (almost) independent of the current sequence number. For example, conditional upon having been inspected once, the probability of a second inspection was 57%; conditional upon having been inspected eight times, the probability of a ninth inspection was 67%.

---

[13] This project used the Fellegi-Sunter technique of record matching, based on the likelihood of agreement on the various fields. For example, exact agreement on establishment name between two inspections makes it very likely that they refer to the same establishment, while agreement on industry (without agreement on several other fields) does not lead to the conclusion that they refer to the same establishment.

The inspections were fairly evenly distributed through time. Approximately 1 in 5 were health inspections. General schedule targeting procedures generated approximately half of the inspections. Complaints or follow-ups to previous inspections motivated approximately one-quarter of the inspections. Accident investigations comprised a minimal 2% of inspections. Inspectors wrote citations in 58% of the inspections, averaging 4 citations across all inspections and 7 citations in inspections with citations.

For the analysis focusing solely on OSHA health performance, we employ two additional datasets, each a subset of the previous one. The second dataset includes all health inspections [N = 63,383]. The third contains only the health inspections with samples of worker exposures to hazardous substances. The descriptive statistics for the health datasets appear in Table A3. Comparing Tables A1 and A3 (col. I), we see that relative to all [health and safety] inspections, health inspections are slightly less likely to have citations (49% relative to 58%) and on average have fewer citations (2.5 relative to 4.2). A greater percentage of health inspections than of safety inspections occur among more intensively inspected plants and larger plants.

Table A3, col. II, indicates that in plants with worker exposure measures, on average, 1.5 exposure samples (27% of the total) violated the standards. Not surprisingly, relative to the average health inspection, inspections with samples have more citations on average, but the difference (3.4 relative to 2.4) is not significant. Otherwise, the two samples look very similar.

4. Empirical Results

The major issue considered in this paper is: Do OSHA's enforcement efforts deter violations of OSHA safety and health standards? In this paper we test the specific-deterrence model of OSHA impact, which posits that repeated inspections of an establishment provide incentives to reduce violations of OSHA standards, and thereby increase workplace safety and health. In the first section below we report on the determinants of health and safety citations in all health and safety inspections. In the second section below, we estimate the model for health inspections only.

We will examine alternative specifications of control variables to determine how sensitive the results are to specification error. The analysis will proceed through increasingly more complete sets of controls, identifying how much additional information the added controls provide as well as the effect they have on the estimated relationships between repeat inspections and compliance.

4.1 Longitudinal Citation Patterns for all Health and Safety Inspections

Table 1 reports the simplest possible longitudinal analysis, identifying how the average number of citations varies with inspection sequence number. The results suggest that the initial inspection of an establishment may reduce subsequent violation levels, but that the following inspections have little effect on compliance. These results are misleading, however, because the agency decision to perform repeated inspections of a plant is highly correlated with poor compliance performance.

Table 2 displays the pattern of violation rates in sequential inspections (SEQNUM), controlling for the total number of plant inspections (NINSP) during the 1972-83 panel period. The pattern, which is remarkably consistent, confirms that the plants OSHA chooses to inspect repeatedly tend to have higher violation rates. To a great extent, the greater the total number of inspections experienced by a plant (NINSP), the greater the number of citations (NUMCITE) for a given inspection sequence number (SEQNUM). The differentiation appears to be weaker among classes with five or more inspections, but it is important to

Table 1. Effect of sequence number on citation variables.
Sample = All safety and health inspections (N=299,295).

SEQNUM	NUMCITE
1	6.3
2	2.8
3	3.1
4	2.7
5	2.8
6	2.6
7	2.6
8	2.6
9	2.5
10-14	2.6
15-19	2.4
20+	2.4

Table 2. Joint effects of SEQNUM and NINSP on citations.
Sample = All safety and health inspections (N=299,295).

A. Mean NUMCITE value.

NINSP	SEQNUM: 1	2	3	4	5	6	7	8	9	10-14	15-19	20+
1	4.2	.	.	.	.	.	.	.	.	.	.	.
2	6.8	1.6	.	.	.	.	.	.	.	.	.	.
3	7.9	2.9	1.9	.	.	.	.	.	.	.	.	.
4	8.7	3.5	3.0	1.9	.	.	.	.	.	.	.	.
5	8.9	4.0	3.4	2.7	2.1	.	.	.	.	.	.	.
6	9.4	4.1	4.1	3.0	2.7	2.1	.	.	.	.	.	.
7	9.6	4.9	4.5	3.2	3.0	2.3	2.2	.	.	.	.	.
8	10.1	5.2	4.5	3.5	3.2	3.0	2.5	2.4	.	.	.	.
9	9.8	4.8	4.9	3.8	3.5	2.6	2.7	2.4	2.3	.	.	.
10-14	10.1	5.3	5.1	3.9	3.6	3.1	2.7	2.5	2.5	2.5	.	.
15-19	10.6	6.4	5.1	3.5	4.5	3.8	3.3	2.9	2.8	2.4	2.4	.
20+	7.2	4.5	5.8	4.7	5.1	4.1	3.8	3.6	2.7	2.9	2.4	2.4

B. Number of inspections.

NINSP	SEQNUM: 1	2	3	4	5	6	7	8	9	10-14	15-19	20+
1	49609	.	.	.	.	.	.	.	.	.	.	.
2	27383	27383	.	.	.	.	.	.	.	.	.	.
3	14179	14179	14179	.	.	.	.	.	.	.	.	.
4	8490	8490	8490	8490	.	.	.	.	.	.	.	.
5	5266	5266	5266	5266	5266	.	.	.	.	.	.	.
6	3231	3231	3231	3231	3231	3231	.	.	.	.	.	.
7	2114	2114	2114	2114	2114	2114	2114	.	.	.	.	.
8	1362	1362	1362	1362	1362	1362	1362	1362	.	.	.	.
9	935	935	935	935	935	935	935	935	935	.	.	.
10-14	1940	1940	1940	1940	1940	1940	1940	1940	1940	4560	.	.
15-19	406	406	406	406	406	406	406	406	406	2030	1017	.
20+	321	321	321	321	321	321	321	321	321	1605	1605	3778

remember that 90% of all plants are inspected 5 or fewer times. Once we have controlled for the total number of inspections of an establishment (NINSP), it is clear that the number of citations declines with repeated inspections (increasing values of SEQNUM). The reduction in citations following the first inspection of a plant is very large: (on average, - 5 citations.)

In order to add controls for additional characteristics of the inspection and of the plant, we perform regression analysis of the determinants of the total number of citations NUMCITE, reported in full in Table A2. In particular we want to test the hypothesis that the reduction in violations across sequential inspections is the same across the groups of plants selected for differential enforcement intensity. In each pair of equations the A version does not include the SEQNUM*NINSP interactions, the B version does include them. Equations 1A and 1B include inspection controls (INORIGIN, INTYPE, INYR); equations 2A and 2B include both inspection and plant controls (SIC, ESTSIZE). For all safety and health inspections, adding plant controls does not reduce the deterrent effect of additional inspections [or substantially change other coefficients] though it increases the adjusted R^2 of the OLS equations from .19 to .20.

Table 3 reports the estimates of the effect of inspections on reducing future citations from Table A2. We report both the interacted and non-interacted estimates. At the bottom of the

Table 3. Effect of past inspections on number of citations by total number of plant inspections (1972-83).

Dependent variable = NUMCITE [mean=4.2, sd=6.7].

Sample = All sample and health inspections [N=299,295].

Total # Inspections	Effect of Inspection #: 1	2	3	4	5+
2	-2.59 (.05)				
3	-2.76 (.07)	-1.05 (.07)			
4	-3.15 (.09)	- .52 (.09)	- .69 (.09)		
5+	-3.23 (.07)	- .47 (.07)	- .52 (.07)	- .32 (.06)	- .028 (.005)
Mean Effect	-2.86 (.04)	- .79 (.04)	- .50 (.05)	- .27 (.05)	- .028 (.005)
F-test pr>F	.000	.000	.146		

Source = Table A3, columns 2A, 2B.

table, we provide the estimated mean effect of that sequence number inspection and the F-test of the hypothesis that the effects estimated for the different values of NINSP are equal.

The average effect of a first inspection in the sample is a reduction of -2.9 citations, with a range of -2.6 to -3.2 citations.[14] The plants that ultimately receive more inspections have the greater reductions. Following a second inspection, the average incremental reduction in citations in this sample is -.8, with a range of -1.1 to -0.5 citations. The pattern of variation across NINSP is reversed in this case, so that the cumulative reductions after two inspections are very close (-3.7 to -3.8 citations) across NINSP sub-groups. The average incremental reductions after 3 and 4 inspections are -.5 and -.3, respectively. Each additional inspection beyond the fourth is associated with an incremental reduction of -.3 to -.4 citations. Though the effects for the first and second inspections are significantly different from each other, the magnitude of the differences is very small. The precision in the estimates arises from the very large sample size [N = 299,295].

It is important to note that these calculations of incremental effects of repeated inspections assume that the reduction in citations induced by an inspection is permanent;

---

[14] We only measure this directly for plants experiencing two or more inspections.

therefore, they provide a conservative estimate of the effect of repeat inspections. Alternatively, if the effect is short-lived, the sum of the coefficients in the SEQNUM series (rather than the value of individual coefficients) represents the full OSHA effect. The former interpretation seems more appropriate when compliance predominantly involves making capital investments with long time horizons; the latter seems more appropriate when compliance primarily involves incurring operating expenditures.

The relevance of this reduction in OSHA violations for the improvement of workplace safety has been challenged by a long succession of authors. For this reason, we are particularly interested in the analysis of health inspections.

4.2 Longitudinal Patterns of Citations and Hazard Exposure Rates in Health Inspections

Table 4 presents data on citation performance in sequential health inspections, controlling for total number of health inspections of an establishment. Health inspections have fewer citations on average than safety inspection, but the pattern of change with additional inspections follows the pattern for safety inspections: citations strongly decline with sequential health inspections, particularly in the first four health inspections of a plant.

We performed regression analysis, controlling separately for past health and safety inspections, total health and total

Table 4. Joint effects of HSEQNUM and HNINSP on citations.
Sample = All health inspections (N=63,383).

A. Mean value: NUMCITE.

HNINSP	HSEQNUM: 1	2	3	4	5	6	7	8	9	10+
1	2.5	.	.	.	.	.	.	.	.	.
2	3.7	1.7	.	.	.	.	.	.	.	.
3	4.0	2.1	1.7	.	.	.	.	.	.	.
4	4.0	2.3	1.7	1.8	.	.	.	.	.	.
5	4.3	2.2	1.8	1.8	1.9	.	.	.	.	.
6	4.4	2.1	2.0	1.5	1.6	2.1	.	.	.	.
7	5.2	2.1	1.8	1.4	1.3	1.3	1.3	.	.	.
8	3.8	2.8	1.5	1.3	1.3	1.3	1.4	2.1	.	.
9	4.3	2.1	2.5	1.2	1.6	0.8	0.9	1.5	2.3	.
10+	3.7	1.9	2.0	1.8	2.3	1.5	2.3	1.0	1.7	1.5

B. Number of inspections.

HNINSP	HSEQNUM: 1	2	3	4	5	6	7	8	9	10+
1	25047	.	.	.	.	.	.	.	.	.
2	6969	6969	.	.	.	.	.	.	.	.
3	2673	2673	2673	.	.	.	.	.	.	.
4	1280	1280	1280	1280	.	.	.	.	.	.
5	686	686	686	686	686	.	.	.	.	.
6	368	368	368	368	368	368	.	.	.	.
7	220	220	220	220	220	220	220	.	.	.
8	149	149	149	149	149	149	149	149	.	.
9	74	74	74	74	74	74	74	74	74	.
10+	173	173	173	173	173	173	173	173	173	666

safety inspections conducted during the panel period, and inspection and plant factors. Table A4 reports the full analysis of the determinants of NUMCIT in health inspections. Table 5 summarizes the estimated deterrent effects of past health inspections on the number of citations. As the sample size declines with the two health sub-samples, the null hypothesis of equal SEQNUM coefficients across NINSP categories is increasingly accepted.

The average number of citations in health inspections during the sample period was 2.5, (s.d.=5.0). The estimated mean incremental effects of the first four inspections were -1.2, -.4, -.1, -.02, with each additional inspection beyond the fourth reducing citations by -.09. The estimated effect of the first inspection ranged from -.9 to -1.3, with no discernible pattern across NINSP sub-groups. The null hypothesis of equality across coefficients is rejected at the 7% level for the first inspection effects. For the second and third inspections, the coefficients are not significantly different from one another. Again, the cumulative effects of first and second inspections vary less across NINSP than the individual inspection effects.

Finally we examine the data on worker exposure measures gathered during health inspections. As noted earlier, 42% of all health inspections report samples. Table 6 reports the pattern of taking samples across health inspection sequence number [HSEQNUM] controlling for total number of health inspections

Table 5. Effect of past health inspections on number of citations, by total number of health inspections (1972-83).

Dependent variable = NUMCITE [mean=2.5, sd=5.0].

Sample = All health inspections [N=63,385].

Total # Health Inspections	Effect of Health Inspection #: 1	2	3	4	5+
2	-1.27				
	(.09)				
3	- .96	- .54			
	(.13)	(.13)			
4	- .85	- .45	- .20		
	(.19)	(.19)	(.19)		
5+	-1.24	- .20	- .18	.034	- .09
	(.107)	(.19)	(.17)	(.150)	(.03)
Mean Effect	-1.16	- .35	- .12	- .019	- .09
	(.07)	(.09)	(.11)	(.130)	(.03)
F-test pr>F	.07	.27	.99		

Source = Table A6, columns 2A, 2B.

Table 6. Joint effects of HSEQNUM and HNINSP on exposure violations.

Sample I = All health samples (N=63,383).

A. Percentage of inspections with samples

HNINSP	HSEQNUM: 1	2	3	4	5	6	7	8	9	10+
1	0.38	.	.	.	.	.	.	.	.	.
2	0.55	0.32	.	.	.	.	.	.	.	.
3	0.64	0.41	0.33	.	.	.	.	.	.	.
4	0.67	0.47	0.38	0.30	.	.	.	.	.	.
5	0.69	0.51	0.45	0.36	0.28	.	.	.	.	.
6	0.73	0.49	0.49	0.38	0.37	0.29	.	.	.	.
7	0.68	0.50	0.50	0.41	0.37	0.26	0.21	.	.	.
8	0.68	0.52	0.44	0.40	0.31	0.30	0.28	0.32	.	.
9	0.84	0.57	0.55	0.47	0.38	0.31	0.30	0.30	0.28	.
10+	0.66	0.55	0.62	0.45	0.48	0.40	0.39	0.35	0.37	0.26

Sample II = All health inspections with samples (N=26,386).

B. Mean value: NUMBAD

HNINSP	HSEQNUM: 1	2	3	4	5	6	7	8	9	10+
1	1.1	.	.	.	.	.	.	.	.	.
2	1.4	1.5	.	.	.	.	.	.	.	.
3	1.8	1.6	1.8	.	.	.	.	.	.	.
4	1.7	1.4	1.9	2.2	.	.	.	.	.	.
5	2.2	1.7	1.7	1.8	2.6	.	.	.	.	.
6	2.0	1.5	2.0	1.9	2.2	3.8	.	.	.	.
7	2.5	1.4	2.0	1.9	2.1	2.8	2.6	.	.	.
8	1.8	2.3	1.9	1.4	1.2	1.4	1.4	.	.	.
9	2.5	1.6	2.3	1.5	1.6	2.6	2.2	3.2	4.4	.
10+	1.6	1.6	1.0	1.6	1.6	1.1	1.4	1.8	1.9	2.3

C. Number of inspections with samples

HNINSP	HSEQNUM: 1	2	3	4	5	6	7	8	9	10+
1	9395	.	.	.	.	.	.	.	.	.
2	3836	2218	.	.	.	.	.	.	.	.
3	1721	1107	869	.	.	.	.	.	.	.
4	853	601	480	383	.	.	.	.	.	.
5	472	353	310	246	.	.	.	.	.	.
6	269	182	180	140	137	108	.	.	.	.
7	149	109	110	91	81	58	46	.	.	.
8	102	78	65	59	46	44	41	47	.	.
9	62	42	41	35	28	23	22	22	21	.
10+	114	96	108	77	83	69	68	60	64	171

[HNINSP]. The percentage of inspections with samples declines with the sequence number, controlling for HNINSP. For a given HSEQNUM, the percentage of plants with samples increases with the total number of health inspections. Focusing only on the plants with samples, we see in panel B of Table 6 that the number of samples in violation does not follow any discernible pattern. With the addition of controls for total safety inspections and for inspection and plant characteristics, however, the familiar pattern emerges in Table 7. With the sample size declining further, the precision is declining, notably for estimates of the effects of inspections following the first one. The effect of the first inspection is on average to reduce the number of overexposures by -.42, with all the estimates close to the average. Only one estimate of the second inspection effect is significant, -.37 overexposures, for the group receiving three inspections. For the third inspection, the average reduction is -.13 overexposures. The estimated effects of the fourth and subsequent inspections are small and very imprecise.

5. Summary and Discussion of the Results

In this paper, we have examined the impact of OSHA enforcement on company compliance with the agency's regulations in the manufacturing sector. Particular emphasis was directed at OSHA's effect in the health area, a relatively unexplored topic in the OSHA literature. We were able to estimate the impact of OSHA enforcement on citations and worker overexposures to

Table 7. Effect of past health inspections on number of worker samples in violation, by total number of health inspections.

Dependent variable = NUMBAD [mean=1.5, sd=3.4].

Sample = All health inspections with samples [N=26,386].

Total # Health Inspections	Effect of Health Inspection #: 1	2	3	4	5+
2	- .42 (.09)				
3	- .38 (.13)	- .37 (.15)			
4	- .44 (.17)	.17 (.20)	- .28 (.22)		
5+	- .48 (.15)	- .03 (.16)	- .22 (.17)	.10 (.16)	- .01 (.04)
Mean Effect	- .42 (.06)	- .13 (.09)	- .13 (.13)	- .06 (.15)	- .01 (.04)
F-test pr>F	.95	.07	.84		

Source = Table A8, columns 2A, 2B.

hazardous substances measured during health inspections. The analysis in Table 5 suggests that in manufacturing plants that ever received a health inspection, OSHA health inspections have reduced the number of citations on average by 28% relative to the number of citations observed during first health inspections.[15]

Of greater interest, perhaps, is the estimate that OSHA health inspections have reduced measured overexposures by 23% among workers in inspected manufacturing establishments where exposure samples were taken.[16] This estimate is simply meant to be suggestive -- the specific figure should not be interpreted literally. It is important to note that our methodology does not allow us to estimate indirect effects of inspections on other non-inspected plants, for example in the same industry or the same geographical region. Also, the analysis is strictly limited to federal OSHA inspections: it does not include the enforcement effects in state plan states. Consequently our estimates of the total effects of OSHA's health initiatives are substantially under-estimated.

These results are particularly noteworthy because this is the first econometric study examining OSHA's impact on overall health quality of regulated establishments. The results contrast sharply with the extensive literature arguing that OSHA has a small (and insignificant) effect on improving workplace safety.

---

[15] We calculated that citations fell by .82 citations from an average of 2.93 citations in first [health] inspections, by applying the effect measured in SEQNUM2 to all plants receiving a first inspection etc.

[16] We calculated a reduction of .30 samples in violation from an average number of documented overexposures in first inspections of 1.33.

The next question is, can we derive any policy implications from the analysis? We are particularly interested in examining OSHA's policy allocating inspection resources. To draw policy inferences, however, we meed to extrapolate beyond the observations in the sample. The results of our tests for sample selection bias provide some justification for making such inferences.

We stated earlier that our sample is a non-random selection of establishments: heavy violators are inspected more intensively. We originally hypothesized that OSHA's may select plants for inspection in part on the basis of characteristics that affect compliance performance, but which are unobservable to us the researchers. As a result, the responsiveness of firms to a given sequence number inspection might vary across the inspection sub-groups experiencing different intensities of enforcement, even after controlling for variations in observable characteristics. We found only weak support for the hypothesis. Systematic differences in responsiveness to enforcement were captured substantially with observable characteristics.[17]

Examining plant responses to inspections, we observe one particularly striking result: a large reduction in violations following the first inspection occurs across all NINSP categories. Even when the estimated effects are significantly different across NINSP, the differences are small and generally decreased when the effects of first and second inspections were combined. Though we have not directly estimated the effect of the first inspection for plants that were never inspected or were once inspected, the similarity of the results across the other groups provides some support for the inference that the effect would be comparable for these establishments.

The large and significant effect of the first inspection in reducing future citations and exposure violations greatly contrasts with the small measured effect of inspections number 5 and beyond (within the 12-year period). The results suggest that on the margin, substantial gains would occur if inspection resources were reallocated from the intensive inspection strategy to the extensive inspection strategy, allowing more first inspections to occur. The analysis suggests that a reallocation from a fifth or greater health inspection to a first health inspection could generate a further reduction in the number of citations by 1.1 (+.12) [mean = 2.5] and in the number of measured worker overexposures by 0.4 (+.3) [mean = 1.5]. This conclusion only applies on the margin: as anticipated future inspection patterns change, firms' responses to current inspections presumably would change. A general equilibrium model of the process, including the generation of expectations of future enforcement activity, is necessary to determine how much reallocation would be optimal.

---

[17] This result is consistent with the requirement imposed by the Barlow decision that the agency must have a clearly defined targetting plan, for the courts to eschew a lengthy procedure showing cause when entry is challenged by the establishment.

REFERENCES

1. Bartel, Ann P. and Lacy Glenn Thomas, "Direct and Indirect Effects of Regulations: A New Look at OSHA's Impact". Journal of Law and Economics. Vol.XXVIII, April 1985. .

2. Cambridge Research Reports, Inc. Public and Worker Attitudes toward Carcinogens and Cancer Risk. Cambridge MA. 1978.

3. Freedman, Audrey. Industry Response to Health Risk. The Conference Board. New York. 1981.

4. Gray, Wayne. "Matching Plants within the OSHA MIS Dataset." Mimeo, Clark University. 1986.

5. Jones, Carol A. et.al. "Methods for Analyzing Compliance with OSHA Health Standards." Final Report to OSHA/DOL. 1986

6. Kochan, Thomas et. al. The Effectiveness of Union-Management Safety Committees. W. E. Upjohn Institute for Employment Research. Kalamazoo MI. 1977.

7. McCaffrey, David, et. al. "Modeling Complexity: Using Dynamic Simulation to Link Regression and Case Studies." Journal of Policy Analysis and Management, Vol. 4, No. 2, 1985.

8. Mendeloff, John. Regulating Safety. MIT Press, 1976.

9. Robertson, Leon S. and J. Phillip Keeve, "Worker Injuries: The Effects of Workers' Compensation and OSHA Inspections." Journal of Health Politics, Policy and Law, Vol. 8, No. 3, Fall 1983.

10. Russell, Louise. "Safety Incentives in Workmen's Compensation Insurance." Journal of Human Resources. Spring 1974.

11. Smith, Robert S. "The Impact of OSHA Inspections on Manufacturing Injury Rates." Journal of Human Resources. Vol. 14, Spring 1979.

12. Viscusi, W. Kip. "The Impact of Occupational Safety and Health Regulations, 1973-1983." Rand Journal of Economics, Winter 1986.

13. Viscusi, W. Kip, "The Impact of Occupational Safety and Health Regulation." Bell Journal of Economics. Spring 1979.

Table A1. Descriptive statistics.

Sample = all safety and health inspections (N=299,295).

Name	Description	Mean	(std. dev.)
NUMCITE	Number of citations on this inspection (includes health and safety citations).	4.15	6.7
SIC2	SIC code (2-digit).	30.36	5.6
SEQNUM	Sequence number of this inspection of this establishment (Dummy variables).		
SEQNUM1	=1 if [Sequence number ≥ 1]	.385	
SEQNUM2	≥ 2]	.219	
SEQNUM3	≥ 3]	.128	
SEQNUM4	≥ 4]	.080	
SEQNUM5	≥ 5]	.188	
SEQNUMC	Continuous variable: =SEQNUM=5 if SEQNUM>5; =0 otherwise.	.841	4.111
NINSP	Number of total inspections of this establishment (Dummy variables)		
NINSP1	=1 if [Total inspections ≥ 1]	.163	
NINSP2	≥ 2]	.183	
NINSP3	≥ 3]	.142	
NINSP4	≥ 4]	.114	
NINSP5	≥ 5]	.398	
NINSPC	Continuous variable: =NINSP-5 if NINSP>5 =0 otherwise.	1.998	6.521
SEQNUM*NINSP	Interactions between inspection Sequence number and total number of inspections		
SEQ2*NINSP2	=1 if SEQNUM2=1 and NINSP2=1]	.09	
SEQ2*NINSP3	SEQNUM2=1 and NINSP3=1]	.10	
SEQ2*NINSP4	SEQNUM2=1 and NINSP4=1]	.09	
SEQ2*NINSP5	SEQNUM2=1 and NINSP5=1]	.34	
SEQ3*NINSP3	SEQNUM3=1 and NINSP3=1]	.05	
SEQ3*NINSP4	SEQNUM3=1 and NINSP4=1]	.06	
SEQ3*NINSP5	SEQNUM3=1 and NINSP5=1]	.29	
SEQ4*NINSP4	SEQNUM4=1 and NINSP4=1]	.03	
SEQ4*NINSP5	SEQNUM4=1 and NINSP5=1]	.24	
SEQ5*NINSP5	SEQNUM5=1 and NINSP5=1]	.19	

Table A1. (continued).

Name	Description	Mean
ACCIDENT	=1 if [Origin of inspection = accident]	.023
COMPLAINT	= complaint]	.220
GENERAL	= general]	.535
FOLLOWUP	= followup]	.222
YR72	=1 if [Year of inspection = 72]	.017
YR73	= 73]	.067
YR74	= 74]	.108
YR75	= 75]	.120
YR76	= 76]	.107
YR77	= 77]	.103
YR78	= 78]	.092
YR79	= 79]	.084
YR80	= 80]	.084
YR81	= 81]	.072
YR82	= 82]	.088
YR83	= 83]	.055
HEALTH	=1 if [Category of inspection = health]	.212
SAFETY	= safety]	.788
ESTSIZE1	=1 if [Number of employees < 20]	.246
ESTSIZE2	= 20-99]	.402
ESTSIZE3	= 100-499]	.251
ESTSIZE4	⩾ 500]	.101

Table A2. Determinants of citations.

Sample = All health and safety inspections (N=299,295).

Dependent variable = NUMCITE (number of citations), [mean=4.15, sd=6.7].

(Standard errors are below coefficients in parentheses)

	1A	1B	2A	2B
CONSTANT	3.40 (.05)	3.41 (.05)	5.18 (.10)	5.21 (.10)
Enforcement				
SEQNUM2	-2.84 (.04)		-2.86 (.04)	
SEQNUM3	-.79 (.04)		-.79 (.04)	
SEQNUM4	-.50 (.05)		-.50 (.05)	
SEQNUM5	-.15 (.05)		-.27 (.05)	
SEQNUMC	-.034 (.004)	-0.34 (.007)	-.028 (.005)	- .028 (.005)
SEQ2*NINSP2		-2.59 (.06)		-2.59 (.06)
SEQ2*NINSP3		-2.74 (.07)		-2.76 (.07)
SEQ2*NINSP4		-3.12 (.09)		-3.15 (.09)
SEQ2*NINSP5		-3.19 (.07)		-3.23 (.07)
SEQ3*NINSP3		-1.08 (.07)		-1.05 (.07)
SEQ3*NINSP4		- .53 (.09)		- .52 (.09)
SEQ3*NINSP5		- .45 (.07)		- .47 (.07)
SEQ4*NINSP4		- .73 (.09)		- .69 (.09)
SEQ4*NINSP5		- .52 (.07)		- .52 (.07)
SEQ5*NINSP5		- .21 (.06)		- .32 (.06)

Table A2. (continued).

	1A	1B	2A	2B
Plant Enforcement Controls				
NINSP2	2.21 (.04)	2.09 (.05)	2.10 (.04)	1.97 (.05)
NINSP3	3.09 (.05)	3.13 (.06)	2.84 (.05)	2.86 (.06)
NINSP4	3.81 (.05)	3.96 (.07)	3.41 (.05)	3.56 (.07)
NINSP5	4.76 (.05)	4.87 (.06)	4.15 (.05)	4.28 (.06)
NINSPC	.011 (.003)	.012 (.003)	-.011 (.003)	-.012 (.003)
Inspection Controls				
HEALTH	-1.87 (.03)	-1.87 (.03)	-1.92 (.03)	-1.92 (.03)
ACCIDENT	-2.98 (.08)	-2.99 (.08)	-3.20 (.08)	-3.20 (.08)
COMPLAINT	- .71 (.03)	- .72 (.03)	- .90 (.03)	- .90 (.03)
FOLLOWUP	-4.85 (.03)	-4.87 (.03)	-4.88 (.03)	-4.89 (.03)
YR72	.86 (.10)	.84 (.10)	.72 (.10)	.71 (.10)
YR73	1.22 (.07)	1.21 (.07)	1.16 (.07)	1.16 (.07)
YR74	1.68 (.06)	1.68 (.06)	1.76 (.06)	1.76 (.06)
YR75	2.44 (.06)	2.43 (.06)	2.59 (.06)	2.59 (.06)
YR76	2.92 (.06)	2.91 (.06)	3.07 (.06)	3.07 (.06)
YR77	1.64 (.06)	1.63 (.06)	1.80 (.06)	1.79 (.06)
YR78	1.41 (.06)	1.40 (.06)	1.63 (.06)	1.62 (.06)

Table A2. (continued).

	1A	1B	2A	2B
Inspection Controls (continued)				
YR79	1.59 (.06)	1.57 (.06)	1.75 (.06)	1.75 (.06)
YR80	1.57 (.06)	1.55 (.06)	1.73 (.06)	1.72 (.06)
YR81	.91 (.06)	.90 (.06)	1.00 (.06)	.99 (.06)
YR82	- .08 (.06)	-.08 (.06)	- .06 (.06)	- .06 (.06)
Plant Controls				
ESTSIZE1			-1.76 (.05)	-1.78 (.05)
ESTSIZE2			- .98 (.05)	-1.00 (.05)
ESTSIZE3			- .36 (.05)	- .38 (.05)
SIC2	No	No	Yes	Yes
R^2 (adjusted)	.186	.186	.196	.196

Table A3. Descriptive statistics for health inspection samples.

Sample I = All health inspections (N=63,383).

Sample II = All health inspections with worker exposure measures (N=26,386).

Name	Description	I Mean (std. dev.)	II Mean (std. dev.)
NUMBAD	Number of worker exposure samples in violation of relevant permissible exposure limit	---	1.486 (3.390)
NUMCITE	Number of citations on this inspection (includes health and safety citations).	2.475 (4.962)	3.433 (5.833)
SIC2	SIC code (2-digit).	31.3 (5.0)	31.2 (5.0)
SSEQNUM	Sequence number of safety inspections of this establishment (Dummy variables).		
SSEQNUM1	=1 if [Sequence number ≥ 1]	.684	.717
SSEQNUM2	≥ 2]	.459	.457
SSEQNUM3	≥ 3]	.298	.282
SSEQNUM4	≥ 4]	.197	.180
SSEQNUM5	≥ 5]	.134	.120
SSEQNUMC	Continuous variable: =SEQNUM-5 if SEQNUM>5; =0 otherwise.	.598 (3.555)	.486 (3.064)
SNINSP	Number of safety inspections of this establishment (Dummy variables)		
SNINSP1	=1 if [Safety inspections ≥ 1]	.168	.162
SNINSP2	≥ 2]	.159	.164
SNINSP3	≥ 3]	.128	.138
SNINSP4	≥ 4]	.097	.104
SNINSP5	≥ 5]	.264	.300
SNINSPC	Continuous variable: =NINSP-5 if NINSP>5 =0 otherwise.	1.260 (5.440)	1.375 (5.566)
HSEQNUM	Sequence number of health inspections of this establishment (Dummy variables)		
HSEQNUM1	=1 if [Sequence number ≥ 1]	1.000	1.000
HSEQNUM2	≥ 2]	.406	.357
HSEQNUM3	≥ 3]	.208	.175
HSEQNUM4	≥ 4]	.119	.093
HSEQNUM5	≥ 5]	.072	
HSEQNUMC	Continuous variable: =HSEQNUMC-5 if HSEQNUM>5 =0 otherwise	.157 (1.017)	.105 (.790)

Table A3. (continued).

Name	Description	I Mean (std. dev.)	II Mean (std. dev.)
HNINSP	Total number of health inspections of this establishment (Dummy variables)		
HNINSP1	=1 if [Total number ≥ 1]	1.000	1.000
HNINSP2	≥ 2]	.220	.224
HNINSP3	≥ 3]	.127	.140
HNINSP4	≥ 4]	.081	.088
HNINSP5	≥ 5]	.178	.187
HNINSPC	Continuous variable: =HNINSP-5 if HNINSP>5 =0 otherwise.	.497 (1.960)	.497 (1.941)
HSEQNUM*HNINSP	Interactions between health inspection number and total health inspections		
HSEQ2*HNIN2	=1 if HSEQNUM=2 and HNINSP=2]	.110	.084
HSEQ2*HNIN3	HSEQNUM=2 and HNINSP=3]	.084	.075
HSEQ2*HNIN4	HSEQNUM=2 and HNINSP=4]	.061	.055
HSEQ2*HNIN5	HSEQNUM=2 and HNINSP=5]	.151	.142
HSEQ3*HNIN3	HSEQNUM=3 and HNINSP=3]	.042	.033
HSEQ3*HNIN4	HSEQNUM=3 and HNINSP=4]	.040	.033
HSEQ3*HNIN5	HSEQNUM=3 and HNINSP=5]	.125	.110
HSEQ4*HNIN4	HSEQNUM=4 and HNINSP=4]	.020	.015
HSEQ4*HNIN5	HSEQNUM=4 and HNINSP=5]	.099	.079
HSEQ5*HNIN5	HSEQNUM=5 and HNINSP=5]	.072	.054
ACCIDENT	=1 if [Origin of inspection = accident]	.008	.003
COMPLAINT	= complaint]	.393	.429
GENERAL	= general]	.424	.443
FOLLOWUP	= followup]	.170	.125
YR72	=1 if [Year of inspection = 72]	.010	.001
YR73	= 73]	.032	.046
YR74	= 74]	.046	.080
YR75	= 75]	.077	.120
YR76	= 76]	.085	.125
YR77	= 77]	.108	.135
YR78	= 78]	.114	.143
YR79	= 79]	.114	.099
YR80	= 80]	.126	.081
YR81	= 81]	.110	.069
YR82	= 82]	.112	.067
YR83	= 83]	.066	.033
ESTSIZE1	=1 if [Number of employees < 20]	.164	.128
ESTSIZE2	= 20-99]	.366	.360
ESTSIZE3	= 100-499]	.319	.351
ESTSIZE4	≥ 500]	.152	.161

Table A4. Determinants of citations in health inspections.
Sample = All health samples (N=63,383).
Dependent variable = NUMCITE [mean=2.48, sd=4.96].
(Standard errors are in parentheses below coefficients)

	1A	1B	2A	2B
CONSTANT	2.99 (.09)	3.01 (.09)	2.83 (.16)	2.86 (.16)
Enforcement				
HSEQNUM2	-1.16 (.07)		-1.16 (.07)	
HSEQNUM3	- .33 (.09)		- .35 (.09)	
HSEQNUM4	- .12 (.09)		- .35 (.09)	
HSEQNUM5	- .01 (.13)		- .02 (.13)	
HSEQNUMC	-.09 (.03)	- .09 (.03)	- .09 (.03)	- .09 (.03)
HSEQ2*HNIN2		-1.30 (.04)		-1.27 (.09)
HSEQ2*HNIN3		- .95 (.14)		- .96 (.13)
HSEQ2*HNIN4		- .82 (.19)		- .85 (.19)
HSEQ2*HNIN5		-1.20 (.17)		-1.24 (.17)
HSEQ3*HNIN3		- .55 (.13)		- .54 (.13)
HSEQ3*HNIN4		- .45 (.19)		- .45 (.19)
HSEQ3*HNIN5		- .17 (.17)		- .20 (.17)
HSEQ4*HNIN4		- .21 (.19)		- .20 (.19)
HSEQ4*HNIN5		- .18 (.17)		- .18 (.17)
HSEQ5*HNIN5		- .05 (.15)		.03 (.15)

Table A4. (continued).

	1A	1B	2A	2B
Enforcement (continued)				
SSEQNUM1	-1.48	-1.48	-1.39	-1.39
	(.07)	(.07)	(.07)	(.07)
SSEQNUM2	- .21	- .21	- .17	- .17
	(.07)	(.07)	(.07)	(.07)
SSEQNUM3	.11	.11	.11	.11
	(.08)	(.08)	(.08)	(.08)
SSEQNUM4	- .11	- .12	- .09	- .09
	(.10)	(.10)	(.10)	(.10)
SSEQNUM5	- .15	- .15	- .15	- .14
	(.10)	(.10)	(.10)	(.10)
SSEQNUMC	.024	.022	.025	.023
	(.013)	(.013)	(.013)	(.014)
Plant Enforcement Controls				
HNINSP2	1.17	1.25	1.16	1.22
	(.06)	(.07)	(.06)	(.07)
HNINSP3	1.69	1.63	1.67	1.61
	(.08)	(.10)	(.08)	(.10)
HNINSP4	1.84	1.68	1.85	1.69
	(.09)	(.14)	(.10)	(.14)
HNINSP5	2.01	1.98	2.00	2.01
	(.10)	(.13)	(.10)	(.13)
HNINSPC	- .009	- .01	- .001	- .004
	(.019)	(.02)	(.019)	(.019)
SNINSP1	.92	.92	.77	.78
	(.08)	(.08)	(.08)	(.08)
SNINSP2	1.06	1.06	.85	.85
	(.09)	(.09)	(.09)	(.09)
SNINSP3	1.17	1.17	.93	.93
	(.10)	(.10)	(.10)	(.10)
SNINSP4	1.14	1.14	.90	.90
	(.11)	(.11)	(.11)	(.11)
SNINSP5	1.14	1.14	.86	.86
	(.11)	(.11)	(.11)	(.11)
SNINSPC	-.001	-.001	-.013	-.012
	(.009)	(.009)	(.009)	(.009)

Table A4. (continued).

	1A	1B	2A	2B
Inspection Controls				
ACCIDENT	- .24	- .25	- .03	- .04
	(.22)	(.22)	(.21)	(.21)
COMPLAINT	- .36	- .36	- .30	- .30
	(.05)	(.05)	(.05)	(.05)
FOLLOWUP	-2.05	-2.05	-2.09	-2.09
	(.07)	(.07)	(.07)	(.07)
YR72	- .24	- .26	- .37	- .38
	(.21)	(.21)	(.22)	(.22)
YR73	- .95	- .97	- .99	-1.00
	(.14)	(.14)	(.14)	(.14)
YR74	- .51	- .53	- .37	- .39
	(.12)	(.12)	(.12)	(.12)
YR75	.29	.27	.42	.40
	(.11)	(.11)	(.11)	(.11)
YR76	- .33	- .35	- .15	- .17
	(.10)	(.11)	(.10)	(.11)
YR77	- .70	- .73	- .53	- .56
	(.10)	(.10)	(.10)	(.10)
YR78	- .39	- .43	- .23	- .25
	(.10)	(.10)	(.10)	(.10)
YR79	- .16	- .19	.002	- .029
	(.10)	(.10)	(.097)	(.098)
YR80	.07	.05	.21	.19
	(.09)	(.10)	(.09)	(.10)
YR81	- .29	- .31	- .21	- .22
	(.10)	(.10)	(.09)	(.09)
YR82	.16	.16	.22	.22
	(.09)	(.09)	(.09)	(.09)
Plant Controls				
ESTSIZE1			.09	.09
			(.08)	(.08)
ESTSIZE2			.33	.32
			(.07)	(.07)
ESTSIZE3			.15	.15
			(.07)	(.07)
SIC2	No	No	Yes	Yes
R^2 (corrected)	.055	.055	.065	.065

Table A5. Determinants of number of samples violating exposure standards in health inspections.

Sample = All health inspections with samples (N=26,386).

Dependent variable = NUMBAD [mean=1.49, sd=3.39].

(Standard errors are in parentheses below coefficients)

	1A	1B	2A	2B
CONSTANT	1.93 (.12)	1.93 (.12	2.69 (.12)	2.69 (.19)
Enforcement				
HSEQNUM2	- .38 (.07)		- .42 (.06)	
HSEQNUM3	- .12 (.09)		- .13 (.09)	
HSEQNUM4	- .15 (.13)		- .13 (.13)	
HSEQNUM5	- .12 (.15)		- .06 (.15)	
HSEQNUMC	- .01 (.04)	- .01 (.04)	- .01 (.04)	- .01 (.04)
HSEQ2*HNIN2		- .38 (.09)		- .42 (.09)
HSEQ2*HNIN3		- .32 (.13)		- .38 (.13)
HSEQ2*HNIN4		- .38 (.10)		- .44 (.17)
HSEQ2*HNIN5		- .44 (.15)		- .48 (.15)
HSEQ3*HNIN3		- .40 (.15)		- .37 (.15)
HSEQ3*HNIN4		- .15 (.20)		- .17 (.20)
HSEQ3*HNIN5		- .02 (.16)		- .03 (.16)
HSEQ4*HNIN4		- .31 (.23)		- .28 (.22)
HSEQ4*HNIN5		- .26 (.17)		- .22 (.17)
HSEQ5*HNIN5		- .15 (.17)		.10 (.16)

Table A5. (continued).

	1A	1B	2A	2B
Enforcement (continued)				
SSEQNUM1	- .04	- .04	- .02	- .02
	(.07)	(.07)	(.07)	(.06)
SSEQNUM2	.14	.14	- .11	- .33
	(.07)	(.07)	(.07)	(.07)
SSEQNUM3	.13	.13	.13	.13
	(.08)	(.08)	(.08)	(.08)
SSEQNUM4	.03	.03	.07	.07
	(.11)	(.11)	(.10)	(.10)
SSEQNUM5	- .14	- .14	- .19	- .19
	(.11)	(.11)	(.11)	(.11)
SSEQNUMC	.026	.030	.031	.033
	(.014)	(.014)	(.014)	(.014)
Plant Enforcement Controls				
HNINSP2	.54	.54	.47	.47
	(.06)	(.06)	(.06)	(.06)
HNINSP3	.91	.94	.75	.78
	(.07)	(.09)	(.07)	(.09)
HNINSP4	.99	.91	.83	.75
	(.09)	(.12)	(.09)	(.12)
HNINSP5	1.39	1.39	1.12	1.14
	(.09)	(.11)	(.09)	(.11)
HNINSPC	- .03	- .028	- .017	- .015
	(.02)	(.019)	(.018)	(.018)
SNINSP1	.17	.19	.03	.03
	(.09)	(.09)	(.08)	(.08)
SNINSP2	.35	.35	.11	.11
	(.09)	(.09)	(.09)	(.09)
SNINSP3	.40	.39	.08	.08
	(.10)	(.10)	(.10)	(.10)
SNINSP4	.50	.50	.12	.12
	(.11)	(.11)	(.11)	(.11)
SNINSP5	.71	.71	.21	.21
	(.10)	(.11)	(.11)	(.11)
SNINSPC	- .003	- .004	- .027	- .027
	(.008)	(.008)	(.008)	(.008)

Table A5. (continued).

	1A	1B	2A	2B
Inspection Controls				
ACCIDENT	- .54 (.36)	- .54 (.36)	- .55 (.35)	- .56 (.35)
COMPLAINT	- .69 (.05)	- .69 (.05)	- .65 (.05)	- .66 (.05)
FOLLOWUP	- .17 (.07)	- .17 (.07)	- .31 (.07)	- .32 (.07)
YR72	.13 (.55)	.12 (.55)	- .47 (.54)	- .48 (.54)
YR73	- .96 (.15)	- .97 (.15)	-1.16 (.15)	-1.17 (.15)
YR74	-1.44 (.14)	-1.45 (.14)	-1.43 (.14)	1.44 (.14)
YR75	-1.43 (.13)	-1.44 (.13)	-1.37 (.13)	-1.38 (.13)
YR76	-1.50 (.13)	-1.51 (.13)	-1.35 (.13)	-1.36 (.13)
YR77	-1.38 (.13)	-1.39 (.13)	-1.27 (.13)	-1.28 (.13)
YR78	-1.33 (.13)	-1.34 (.13)	-1.22 (.13)	-1.23 (.13)
YR79	- .34 (.13)	- .35 (.13)	- .23 (.13)	- .24 (.13)
YR80	.23 (.13)	.22 (.13)	.29 (.13)	.29 (.13)
YR81	.17 (.14)	.16 (.14)	.20 (.13)	.19 (.13)
YR82	.24 (.14)	.24 (.14)	.24 (.13)	.24 (.13)
Plant Controls				
ESTSIZE1			-1.17 (.09)	-1.16 (.09)
ESTSIZE2			- .81 (.07)	- .81 (.07)
ESTSIZE3			- .31 (.07)	- .31 (.07)
SIC2	No	No	Yes	Yes
R^2 (adjusted)	.062	.062	.102	.102

HITTING THE "T":
THE REGULATORY LOSS OF SOIL MANAGEMENT POLICY*

Robert E. Rosenman
Department of Economics
Washington State University
Pullman, WA 99164-4860

June, 1987

PRELIMINARY

---

* This material is based upon work supported by the National Science Foundation under Grant No. SES-8605052. Rod Fort provided helpful comments. Charles Whiteman corrected several errors in the proof of Proposition 2, and offered suggestions which improved this paper. Any remaining errors are my own.

Hitting the "T":

The Regulatory Loss of Soil Management Policy

ABSTRACT

The federal government has in its active role of encouraging soil conservation taken as a goal the "T", or tolerance level, of the soil being managed. However, the success of the hitting the "T" has been limited. This paper explains the basic framework of a dynamic regulatory "game" between the government and the farmers it is trying to influence, and explores how the way farmers react to interventionist activities may be responsible for the limited success in encouraging soil conservation. Within this framework new techniques for formulating effective policies are explored, and an empirical application to Maine potato farmers is performed.

I. Introduction

The federal government, primarily through the Soil Conservation Service (SCS) and the Agricultural Conservation Program (ACP)[1], has long been an active participant in the conservation of soil resources. Its programs have tried to encourage farmers to increase the soil management that they practice. The involvement of the government has focused on the goal of the "T", or tolerance level of the soil being managed. The "T" is that rate of erosion which does not affect the long term productivity of the soil. Economists have argued long and hard that the "T" is not the proper level of soil management from a welfare analysis perspective.[2] However, our ability to influence policymakers has been limited. Thus, since the political process has evolved to the goal of "T" level of erosion, it may be interpreted in some sense as a regulatory objective. As with any goal, the policymakers face tradeoffs with budgetary considerations and other programs in meeting this objective.

Within this constraint there are optimal and suboptimal ways of formulating a policy intervention to move farmers towards the objective of "T" erosion. The difference comes from how the policymaker interacts with the farmers. Because of the dynamic elements in a soil management decision, the policymakers (essentially the SCS and ACP) play a dynamic game with the farmers, and how the farmers respond limits the effectiveness of policy rules. (This point was made clear in Lucas's (1976) critique of econometric policy evaluation). For the policymakers, this means any rule changes the future expectations farmers formulate, and thus they may not respond as expected. Upon reevaluating the policy, a new rule will be established. As the process converges, the policy rule may evolve to a time consistent one upon which the policymaker cannot gain by revising.

However, Whiteman (1986) shows that a rule which dominates the time consistent one is a precommitment[3] strategy: at the beginning of the "game" the policymaker sets a rule and promises never to change it. For this rule he exploits the fact that expectations respond to the rule, and maximizes his objective (in this case hitting the "T") subject to that constraint. The contention of this paper is that the ASCS (with its ACP) followed the former policy (in pursuit of short term gains), evolving to a time consistent rule, which resulted in a larger deviation from the "T" than if the optimal precommitment strategy had been followed. That is, in formulating policies for soil management subsidization the ASCS neglected the effects its own decisions have on farmers expectations of future soil management decisions. This paper argues that these effects are important and thus the government should have viewed the policy problem as a game against intelligent agents.

Many researchers have recognized the impotence of soil conservation policies and have tried to analyze reasons why. However, most studies have been suggestive or discursive rather than analytical. Batie (1982), for example, reviews policies that support and encourage soil conservation, with a discussion of how different policies both help and hurt. Similarly, Easter and Cotner (1982) conclude that current policies are ineffective, but also add that the data needed to evaluate the strategies are not available.

A more analytical study was done by Forster and Becker (1979). It contrasts three alternative policies; restrictions on soil loss, taxes on soil loss, and subsidies for erosion control. In an application to the Lake Erie watershed basin, the different policies are shown to have nearly the same net economic benefits, although the distribution of these benefits differ.

Given the ineffectiveness of current policies, many researchers have concluded that we need a better idea of why farmers adopt (or don't adopt) soil management practices. For example, Napier, et al (1984, p.205), argue that since current policies are ineffective "... new mechanisms must be developed. Farmers must be motivated to adopt conservation practices without 'bribing' them with monetary incentives or 'forcing' them to correct the problem." They conclude that our focus as researchers should be to identify factors that help predict why farmers choose soil conservation practices.*

This paper takes an alternative view. Rather than try to explain why farmers adopt soil management practices, and then design new programs to exploit that understanding, I follow the lead of Moore, et al (1979) and Prato (1987). These papers looks at current programs which share the cost of soil management and analyze how they can be reorganized to make soil management practices financially attractive to farmers (Moore) or more effective (Prato). Herein is the problem which is the focus of this paper. How to take an (existing) policy (in my particular application the ACP) and make it more effective in achieving its goals.

To relate this point to the earlier discussion of optimal and consistent plans, a very brief description of the ACP would be useful.[8] The ACP is a cost sharing program for the adoption of soil conservation practices. Although until 1978 there was a cap of $2,500.00 per individual per year, there has always been considerable leeway in cost share rates and eligibility. They are determined at the county level by farmer representatives (within federal and state guidelines) and these representatives decide who receives funds. This discretion is not always beneficial. Rasmussen (1982) points out that much of the ACP dollars spent prior to 1980 went for production-oriented practices rather than more focused soil conservation. My contention is that much of the failing in the ACP may be due to the discretionary orientation of the program. While this may make it more attractive to farmers in the onset, in the long run the dynamic interaction leads to policy ineffectiveness.

The remainder of this paper is organized as follows. In section II the problem facing the policymaker is explained within the constructs of a simple model of farmer decision making. A solution to the farmers' problem under rational expectations is characterized and the policymaker's problem is solved. Section III is an application to Maine potato farmers. and the actual policy values are compared to the optimal ones. Section IV provides some concluding statements. There are three appendices to this paper. The first contains some mathematical derivations used at points in the analysis, the second is the proof of proposition 2, and the third is a short discussion of the data on Maine potato farmers that is used in the application.

II. Soil Management Policy and the Dynamic Game

The SCS and ASCS have used as a goal for soil policy to hit the "T", that is the tolerance level of erosion that the soil can support. This is a "steady state" concept that maintains the soil depth and its long-term productivity.[6] In the soils literature soil loss is usually measured in tons per acre, as given by the Universal Soil Loss Equation (USLE).[7] The USLE is given by T = CxPxRxKxLxS where T is the long-term soil loss (in tons per acre)[8], C and P are measures of cultivation and rotation practices, and the remaining parameters measure the inherent characteristics of the soil at risk, which are essentially beyond the control of the farmer. Heimlich and Bills suggest dividing the USLE into two parts, the soil management measure (CP) and the soil constraints measure (RKLS), which will be termed potential erosion here.

It is now possible to state the policymakers' problem: At each point in time the farmers choose soil management (essentially the CP factors of the USLE) which determines the soil loss when combined with the potential erosion. The policymaker wants to influence the farmer so that the choice of soil management results in a soil loss equal to the tolerance level of the field. Let T* be the tolerance level of the field in question. Then, letting potential erosion (POT) be constant through time, the actual soil loss at time t is

$$T_t = M_t POT \tag{1}$$

where M_t is the soil management choice (the CP factors of the USLE) the farmer makes. The policymaker wants $T_t=T*$, or equivalently, $M_t=M*=T*/POT$. The policymaker tries to alter the farmer's behavior by providing a subsidy, S_t, per unit of M_t practiced. However, the program faces a budget constraint. Therefore, the objective is to hit the T while keeping the variance from the budget to a minimum. If the budgeted amount for soil conservation policies is $S*_t$, an expression which captures the policymaker's objective is to minimize

$$E((M_t-M*)^2 + h(S_t-S*_t)^2) \tag{2}$$

where E is expectations and h is a nonnegative parameter.

Farmers face the problem of choosing soil management practices and inputs to maximize profits. Inputs combine with soil to produce the crop, and the marginal productivity of each is affected by soil management. The primary soil management technique used in this study is crop rotation, so it is the previous year's soil management that effects this year's yield. Thus, the farmer's output, Y_t, is given by

$$Y_t = f(X_t,\ D_t,\ M_{t-1}) \tag{3}$$

where X_t are the inputs used, D_t is the soil depth, M_{t-1} is the previous year's soil management and $f(.,.,.)$ is the production function.[*]

The farmer chooses X and M to maximize the present value of future profits

$$E_{-1}\sum_{t=0}^{\infty} b^t\{Y_t - R_{t-1}X_t + S_{t-1}M_t\} \tag{4}$$

where E_j is conditional expectations at time j, R_t is the real price of inputs (the price of the crop is used as a numeraire) and b is the farmer's rate of time preference (discount factor). Notice, in this specification the farmer purchases the inputs the period prior to there use and knows the subsidy amount prior to investing in soil management practices. Equation (4) assumes there is no direct cost of soil management practices. Since in the data used for our analysis the primary practice is crop rotation, we are implicitly assuming that the alternative crop is break-even in terms of cost and revenue.

The maximization is performed ignoring the fact that

(5) $$D_{t+1} = d_1 D_t + d_2 M_t$$

where d_1 and d_2 are positive parameters. This means farmers account for only the productivity effects of soil management, not the long term effects. We assume that the supply curve for inputs is perfectly elastic (horizontal) except for shocks, and farmers know that

(6) $$R_t = A^*(L)u_t$$

where $A^*(L)$ is a square-summable polynomial in the lag operator (L) and u_t is fundamental for R_t, that is, u_t is the sequence of one-step-ahead forecast errors made from predicting R_t from its own past.

As shown in appendix A, the Euler equations for the farmer's problem can be solved for a decision rule for choosing soil management practices that follows the expectational difference equation

(7) $$E_t M_{t+1} - d_1 E_{t-1} M_t - K_1 d_2 M_t = (1-d_1 L)\{K_2 E_t R_{t+1} + K_3 S_t + K_0\}$$

where the K_i are parameters which are functions of the parameters of the production function. The policymaker seeks to set S_t according to the rule

(8) $$S_t = F^*(L)u_t$$

where $F^*(L)$ is a polynomial in nonnegative powers of L so that the solution to (7) will minimize the value of the objective function. Whiteman (1986) shows the solution to the policymaker's problem when $d_1=0$. However, as we shall see, with feedback in the system (essentially from equation (5)) the solution to the policymaker's problem is somewhat more difficult.

Proposition 1: (Equilibrium Soil Management Rules). When the policymaker precommits to a rule for setting S_t by $S_t = F*(L)u_t$, soil management practices followed by farmers will follow the path $M_t = C(L)u_t + C*$ where

$$C* = [(1-d_1)/(1-p)]K_o \tag{9a}$$

and

$$C(L) = \frac{(1-d_1L)\{LA(L) + LF(L) + C_o\}}{(1-pL)} \tag{9b}$$

where $p = d_1 + K_1d_2$, $C_o = -p^{-1}\{A(p^{-1}) + F(p^{-1})\}$, $A(L) = K_2L^{-1}[A*(L) - A*_o]$ and $F(L) = K_3F*(L)$, as long as $|p| > 1$.

Proof: Using Wiener-Kolmogorov prediction formulas $K_2E_tR_{t+1} = K_1L^{-1}\{A*(L) - A*_o\}u_t = A(L)u_t$. Similarly, supposing that $M_t = C(L)u_t + C*$ gives that $E_tM_{t+1} = L^{-1}\{C(L) - C_o\}u_t + C*$ and $E_{t-1}M_t = L\{E_tM_{t+1}\} = \{C(L) - C_o\}u_t + C*$. By substitution, equation (7) may be written

$$\{L^{-1}[C(L) - C_o] - d_1[C(L) - C_o] - K_1d_2C(L)\}u_t + (1 - d_1 - K_1d_2)C* \tag{7'}$$
$$= (1 - d_1L)\{A(L) + F(L)\}u_t + (1 - d_1)K_o.$$

Let $d_1 + K_1d_2 = p$. Then equating coefficients of the polynomials in L from (7'), after some rearranging gives

$$C* = [(1-d_1)/(1-p)]K_o$$

and

$$C(L) = \frac{(1-d_1L)\{LA(L) + LF(L) + C_o\}}{(1-pL)} .$$

As shown in Whiteman (1983, p.7), stationarity requires that

$$C_o = -p^{-1}\{A(p^{-1}) + F(p^{-1})\}.$$

This completes the proof.

The solution to the policymaker's problem is found by using (9a) and (9b) to calculate the variance of M_t around $M*$ and (8) to compute the variance of S_t around $S*_t$, and then choosing F(L) to minimize expression (2). The methods are an adaptation of those given in Whiteman (1986). Here, it is necessary to formulate the rule within the feedback constraint imposed by (5). We seek the "open loop" form of the policy where policymakers react to the shocks, not the soil management rules. As long as we assume that the farmers behave as if they cannot influence the policy rules, the "closed loop" policies ($S_t = b(L)S_t + B(L)M_t$) offers no strategic advantage (although it may be easier for implementation, since the policymaker need see only the farmers soil management practices, not the shocks to the system). Additionally, under this assumption, the closed loop policy is easily derived from the open loop policy, (as is done in the application in section III).

The derivation of the optimal precommitment open loop policy rule follows the derivation of Theorem 1 in Whiteman (1986).

Proposition 2: (The Optimal Precommitment Soil Management

Policy Rule). Assume that a policy rule of the form (8) is to be employed to minimize the objective function given by (2), subject to the constraints (6) and (7). The z-transform of the coefficients F_0, F_1, ... in the optimal precommitment policy rule is given by

$$(10)\quad F(z) = \frac{(hp+d_1)^{-1}\{zd(z)d(z^{-1})A(z)-md(m)d(m^{-1})A(m)\}}{(1-mz)(1-m^{-1}z)}$$

where $d(z) = 1-d_1z$, $|m|<1$ comes from the factorization

$$g(1-mz)(1-mz^{-1}) = [h(1-pz)(1-pz^{-1}) + (1-d_1z)(1-d_1z^{-1})]$$

and $F*(z) = F(z)/K_3d_2$.[10]

Proof: This proof follows that of Whiteman's Theorem 1 (1986, p.1392). As in his proof we shift to the "frequency domain" using the inversion formula from the covariance generating function for M_t. However, the feedback in the soil transition equation makes this a nontrivial adaptation of Whiteman's proof. For expositional convenience, the details are given in Appendix B.

If h=0 (that is if the ASCS is concerned only with hitting the T) it is easily shown that $F(z)=-A(z)$.[11] Under such a circumstance the ASCS should react to changes in relative prices (that is, changes in R_t) with changes in the subsidy (S_t) to just offset the former and move the farmers to M*. In the more general case the optimal precommitment policy is

$$(11)\quad S_t = mS_{t-1} + (hp+d_1)^{-1}\frac{Ld(L)d(L^{-1})A(L)-md(m)d(m^{-1})A(m)}{1-m^{-1}L}u_t.$$

Substituting (10) into (9b) describes the path of soil management if policymakers follow this rule. When the soil management subsidy is set by the rule given by (11), soil management follows the path[12]

$$(12)\qquad M_t = \frac{(1-d_1L)[hp(1-p^{-1}L)(1-pL)][LA(L)-mA(m)]}{(1-pL)(hp-d_1)(1-mL)(1-m^{-1}L)}u_t$$

and the closed loop precommitment policy is given by

$$(13)\quad S_t = p^{-1}S_{t-1} + \frac{(1+d_1{}^2)}{hp(1-d_1L)}M_t - d_1\frac{(1+L^2)A(L)-(1+m^2)A(m)}{hp(1-d_1L)[LA(L)-mA(m)]}M_t.$$

An interesting difference between equation (13) and the closed loop precommitment policy derived by Whiteman in his generic case is persistance of past soil management on current policy. In fact, the entire history of soil management affects current policy. This is a consequence of the feedback from soil management into the state of the environment (that is, the fact that d_1 does not equal zero).[13] In the next section we derive this rule for subsidizing soil conservation (ACP grants) for Maine potato farmers.

III. The Loss of Soil Management Subsidies: An Application

Tables 1 to 6 contain the estimates of the equations of this model.[14] The first preference was to estimate the production function (equation A.1) and the transition equation for soil, and then use those estimates to derive the other relationships in the model. However, severe multicollinearity in the inputs and soil management data[15] made it impossible to obtain reliable estimates of the parameters of the production function. To overcome this problem the model was estimated at the level of the first-order necessary conditions (Euler equations) for the optimization of the farmer's problem (that is, manipulations of equations A.2 and A.3).

Consider first the derivation from the first order condition with respect to inputs (table 1). This can be construed as the derived demand curve for inputs. The estimates indicate that input demand is positively related to soil management lagged one period (SM in the table) and soil depth (SOILD in the table). Recall that soil management has a one-year lagged effect. Thus, farmers who enhance their soils productivity (by rotation) complement the productivity effect by using more inputs. The same holds for soil depth. Surprisingly, there was also a positive relationship between real relative input prices (INPRI) and inputs used. The estimated equation explains the majority of the variance in input usage, and the marginal significance of 0.87 for the Box-

Pierce Q statistic suggests that there is little evidence of residual correlation.

Estimates of the Euler equation that comes from the first order condition with respect to soil management are reported in table 2A. As before, we have the complementary effect between soil management and inputs. The estimates also show a negative relationship between soil management and soil depth the next period. Farmers compensate for shallow soil by increasing soil management. As would be expected, government subsidies for soil management (ACP) increase the demand for it. Although these estimates explain over 80% of the variation in soil management there is marginal evidence of serial correlation in the errors. Table 2B corrects for the serially correlated errors. All signs are maintained, and the explanatory power of the equation is increased.

Of the three structural equations estimated, the soil transition equation (table 3A) is the least satisfactory. While the estimates of d_1 and d_2 are reasonable, there is strong evidence of serial correlation in the errors of this equation. The serial correlation is corrected in table 3B. Again, there is improvement in the estimates from this correction.

In pursuing the time consistent strategy in attempting to minimize (2), the ASCS presumably "looked at everything." Thus, an estimate of the policy rule followed by the Service during the sample can be obtained by regressing S_t (ACP) on all of the variables that appear in the structural equations and were available at the time the subsidy was set. This estimate is given in table 4. It explains virtually all of the variation in S_t, and the remaining errors are not serially correlated. This estimation is more useful for forecasting than understanding the structural behavior of the ASCS.

An alternative estimate is given in table 5A, which is the consistent decision rule for subsidies which is optimal for minimizing (2), assuming $S^*_t = S_{t-1}$.[16] There is strong evidence of serially correlated errors[17]. In Table 5B the serial correlation is corrected. The estimates in table 5B, along with those from tables 1 to 3, imply that h=2.065.

The final equation for which an estimate is needed is (6), that is $R_t=A^*(L)u_t$. Three procedures were used for analyzing this relationship. First, autoregressive models of several lags were estimated using OLS. Only one of these estimations, an AR(1) with no constant produced a significant estimate for a coefficient. However, this equation had a poor fit. A chi-square test (table 6A)[18] found a significance level of 0.93, indicating no serial correlation. All indications were that R_t follows a white noise. To check this, as reported in table 6B, autocorrelations for six lags were computed. Using the test statistic

$$z = \frac{(n-3)^{.5}}{2} \ln \frac{(1+r_k)(1-r^*_k)}{(1-r_k)(1+r^*_k)}$$

which is approximately normal (see Freund, p.381), the estimated correlation r_k was tested against the value $r^*_k=0$ for $k = 1,2,\ldots 6$. For all k the null hypothesis that $r_k=0$ could not be rejected, and therefore we take R_t to be white noise, that is, $R_t=u_t+R^*$, where R^* is the central (mean) value and $A^*(L)=1$.[19]

Using the estimates discussed above it is easily derived by substitution that (in equation A.6) $K_1=.201$, $K_2=0.001$, $K_3=.412$, $p=1.002$ and $m=0.963$. It is now possible to derive the optimal precommitment rule, using these values, $h=2.065$ and $A(z)=1$ in (11) or (13). Then it will be possible to compare the actual SCS subsidy rules to the optimal one, and derive measures of regulatory loss.

Two measures are used to compare the optimal system to the actual one. First, we examine how each system responds to shocks. To do this, the covariance structure estimated in the system equations was imposed on the optimal and estimated models and we calculated the response of each system to shocks in soil management, soil depth, and government subsidy (ACP). These are presented in Figures 1 through 3. Each figure shows the movement of soil management, soil depth and subsidy under the optimal rule and the actual rule. Figure 1 shows system responses to a shock in soil management, figure 2 is for a shock in soil depth, and figure three is for a shock in ACP (subsidy).

Generally, both systems show the same patterns of responses to shocks, although the optimal system's responses are stronger. (In all plots the first two periods are eliminated for better scaling of the figure). When soil management suffers a (positive) shock (Figure 1) ACP and soil management move towards the initial values (but show a persistent but seemingly stable increase) while soil depth in both systems steadily increases. However, the optimal system moves back towards the initial values faster (for ACP and soil management) and has a stronger increase in soil depth.

A similar pattern holds for shocks to soil depth (Figure 2), although here the effect on soil depth is the about the same for the optimal system and the actual system. The actual system shows a stronger continued growth in soil depth than the optimal one, but also has a stronger persistance in the increase of ACP and soil management.

The real difference is in the response of the systems to a shock in ACP (Figure 3). While there are initial fluctuations in both systems, the optimal system moves towards the initial values, and stabilizes at slight increases in all variables. For the actual system recovery is not as complete for soil management and ACP, and there is a strong (positive) response of soil depth.[20]

A second comparison was made using simulations of the two models. The paths were simulated by using the actual values of INPRI, and solving for the endogenous variables in the models. These are reported in tables 7 through 9 and figures 4 to 7.

For soil management and ACP, the optimal system was somewhat more volatile than the actual one, and although the patterns were similar. Soil depth showed strong persistence. Soil management, ACP and soil depth were also consistently higher for the optimal system.

Of primary interest are the simulated values for soil management. Discussions with SCS staff indicate that a potato, oats rotation (yearly) is considered optimal for the area generating this data (giving an M*=2). We see that the optimal rule fluctuates around a mean of 2.2, while the actual rule was centered around 1.5 (about a potato, potato, oats rotation). This means under the optimal rule, the average soil management would exceed the "T" value, causing an increase in soil depth, as shown in figure 6.

From a regulatory perspective, the key value is that of the objective function. These are given in table 9 and figure 7. A lower value is better and, most interestingly, the optimal rule is not completely dominant over the actual rule in meeting the policymaker's objectives. The values of the objective function were initially greater under the optimal rule, although for remaining 17 of the 23 years calculated, the optimal rule was better. It is clear that there are opportunities for policymakers to extract large short-term gains by using something other than the optimal rule. Initially, the actual rule obtained large benefits. But as the farmers caught on, they changed their expectations, and the value of the objective function increased. By the sixth period the optimal rule becomes better, and remains so. What has happened is that the system evolved to a consistent but not optimal rule. Overall and on average, the optimal rule is better. The average value of the objective function with the optimal rule was 0.252899, while with the actual rule the average value was 0.284950, an improvement of about 13 percent.

IV. Conclusions

The question remains that if the proposed rule is so good, why doesn't the SCS follow something like it? The argument here is that the decision makers are pursuing short-term benefits at the expense of long-term gains. Even if the ASCS promises not to do so in the future, intelligent farmers see the repeated series of actions, and formulate expectations based on the short-term policies. The policymaker responds to the farmers, and the system evolves. By not precommiting at the beginning, the ASCS has relinquished its ability to exploit expectations. This is not because it is stupid, but probably because it operates in a (political) short-run. The appeal of short-term benefits from its policy actions is too great. Essentially, the problem is discretion, and the solution is to remove it.

There are, of course, alternative explanations for why the ASCS does not follow the optimal rule. One which is foremost is that the objective function used in this analysis is not the correct one. That is, despite its rhetoric, the goal of government policymakers may not be to hit the T. If so, then two parts of this paper are of primary importance: first that the ASCS must understand how farmers' reactions and expectations matter (whatever the objective), and second that the optimal rule for achieving "T" erosion can be used as a norm of physical costs, to differentiate from economic costs of soil loss. But if the stated objectives ("T" erosion) of the program are not those actually pursued then complete policy analysis requires a model of ASCS rent seeking, something that lies beyond the bounds of this paper. Here we provide the best rule for a specific goal, hitting the "T", and in any case the analysis gives measures of how actual ASCS behavior misses it.

APPENDIX A

MATHEMATICAL DERIVATIONS

I. Derivation of Equation (7):

Let

(A.1) $$Y_t = a_0 + a_1D_t + a_2X_t + a_3M_{t-1} + a_4X_t^2 + a_5D_t^2 + a_6M_{t-1}^2 + a_7X_tM_{t-1} + a_8X_tD_t + a_9D_tM_{t-1}.$$

This is taken as a quadratic (Taylor) approximation of the true production function. Soil management, that is crop rotation and/or fallow land, has a one year lag on productivity. After substituting (A.1) into (4) the first order necessary conditions for the farmer's optimization problem (ignoring (5)) are

(A.2) $$E_{t-1}[a_2 + 2a_4X_t + a_7M_{t-1} + a_8D_t - R_{t-1}] = 0$$

from the partial derivative with respect to X_t and

(A.3) $$E_{t-1}[ba_3 + 2ba_6M_t + ba_7X_{t+1} + ba_9(d_1D_t + d_2M_t) + S_{t-1}] = 0$$

from the partial derivative with respect to M_t, using the definition of D_{t+1}. Equation (A.2) implies that

(A.4) $$E_{t-1}X_{t+1} = E_{t-1}(1/2a_4)[R_t - a_2 - a_7M_t - a_8(d_1D_t + d_2M_t)]$$

again using the definition of D_{t+1}. Substituting (A.4) into (A.3) and rearranging gives

(A.5) $$G_0 + G_1E_{t-1}M_t + G_2E_{t-1}D_t + G_3E_{t-1}R_t + S_{t-1} = 0$$

where

$G_0 = ba_3 - [ba_2a_7/2a_4]$

$G_1 = 2ba_6 - [(ba_7/2a_4)(a_7 + a_8d_2)] + ba_9d_2$

$G_2 = ba_9d_1 - [ba_7a_8d_1/2a_4]$

$G_3 = ba_7/2a_4$

or

(A.6) $E_{t-1}M_t - K_1E_{t-1}D_t = K_2E_{t-1}R_t + K_3S_{t-1} + K_0$

where

$K_1 = -G_2/G_1$

$K_2 = -G_3/G_1$

$K_3 = -1/G_1$

$K_0 = -G_0/G_1$.

Using equation (5) in the text we can write

(A.7) $E_{t-1}D_t = (1-d_1L)^{-1}d_2M_{t-1}$.

Substituting (A.7) into (A.6), multiplying both sides by $(1-d_1L)$ and leading the result one period gives (7).

II. Derivation of Equation (12):

Substitute (10) into (9b) to get

$$c(z) = \left\{\frac{d(z)}{1-pz}\right\} \Big\{zA(z) + z(hp+d_1)^{-1}(1-mz)^{-1}(1-m^{-1}z)^{-1}zd(z)d(z^{-1})A(z)$$

$$- z(hp+d_1)^{-1}(1-mz)^{-1}(1-m^{-1}z)^{-1}md(m)d(m^{-1})A(m) - p^{-1}A(p^{-1})$$

$$- p^{-1}(hp+d_1)^{-1}(1-mp^{-1})^{-1}(1-m^{-1}p^{-1})^{-1}p^{-1}d(p^{-1})d(p)A(p^{-1})$$

$$+ p^{-1}(hp+d_1)^{-1}(1-mp^{-1})^{-1}(1-m^{-1}p^{-1})^{-1}md(m)d(m^{-1})A(m)\Big\}.$$

Noting that

$$-(hp+d_1)^{-1}(1-mp^{-1})^{-1}(1-m^{-1}p^{-1})^{-1}p^{-1}d(p^{-1})d(p) = 1$$

and so

$$p(hp+d_1)(1-mp^{-1})(1-m^{-1}p^{-1}) = -d(p)d(p^{-1})$$

(using the fact that $[m+m^{-1}] = [h+hp^2+1+d_1^2]/[hp+d_1]$) we get that

$$\text{(A.8)}\quad c(z) = \left[\frac{d(z)}{1-pz}\right]\Bigg[\left\{\frac{1}{d(z)d(z^{-1})} + \frac{z}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\}d(z)d(z^{-1})zA(z)$$

$$+ \left\{\frac{1}{d(p)d(p^{-1})} + \frac{z}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\} d(m)d(m^{-1})mA(m)\Bigg].$$

The first term in the braces of (A.8) can be written

$$\left\{\frac{1}{d(z)d(z^{-1})} + \frac{z}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\} = \frac{(hp+d_1)(1-mz)(1-m^{-1}z)+zd(z)d(z^{-1})}{d(z)d(z^{-1})(hp+d_1)(1-mz)(1-m^{-1}z)}.$$

The numerator reduces to $hp(1-p^{-1}z)(1-pz)$, again using the relationship between $m+m^{-1}$ and h, d_1 and p. Similarly, the second term in braces can be written

$$\frac{hp(1-mz)(1-m^{-1}z) + z\, d(m)d(m^{-1})}{(hp+d_1)(1-mz)(1-m^{-1}z)d(m)d(m^{-1})}$$

using the fact that $d(p)d(p^{-1}) = [(hp+d_1)/hp]d(m)d(m^{-1})$. Again the numerator reduces to $hp(1-p^{-1}z)(1-pz)$. Therefore

$$c(z) = \left\{\frac{d(z)}{1-pz}\right\} \left\{\frac{hp(1-p^{-1}z)(1-pz)zA(z)}{(hp+d_1)(1-mz)(1-m^{-1}z)} + \frac{hp(1-p^{-1}z)(1-pz)mA(m)}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\}$$

or

$$\text{(A.9)}\quad c(z) = \left\{\frac{d(z)}{1-pz}\right\} \left\{\frac{hp(1-p^{-1}z)(1-pz)}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\} \{zA(z) - mA(m)\}$$

III. Derivation of Equation (13):

We seek $S_t = B(L)M_t$. Note that $M_t = C(L)u_t$ and $S_t = F(L)u_t$. Therefore $S_t = F(L)C(L)^{-1}M_t$ or $B(z) = F(z)C(z)^{-1}$. Thus

$$B(z) = \left\{\frac{zd(z)d(z^{-1})A(z)-md(m)d(m^{-1})A(m)}{(hp+d_1)(1-mz)(1-m^{-1}z)}\right\} C(z)^{-1}.$$

Using (A.9) this becomes

$$\text{(A.10)}\quad B(z) = \left\{\frac{(hp)^{-1}}{d(z)(1-p^{-1}z)}\right\} \left\{\frac{zd(z)d(z^{-1})A(z)-md(m)d(m^{-1})A(m)}{zA(z) - mA(m)}\right\}.$$

The desired result is obtained by multiplying $S_t = B(L)m_t$ through by $(1-p^{-1}L)$ and expanding the term in the second set of braces in (A.10).

APPENDIX B

Proof of Proposition 2:

When $M_t = C(L)u_t$, the covariance generating function is given by $g_y(z) = C(z)C(z^{-1})$ where $C(z)$ is the z-transform of the coefficients in $C(L)$. The covariance EM_tM_{t-j} is given by the coefficient on z^j in $g_y(z)$. By using the inversion formula

$$EM_tM_{t-j} = \frac{1}{2\pi i}\oint C(z)C(z^{-1})z^{-j-1}dz$$

where $\oint$ represents contour integration about the unit circle $|z| = 1$. Therefore, the value of the policymakers' objective function (equation 2) in frequency domain notation is

$$J = \frac{1}{2\pi i}\oint \frac{(1-d_1z)[zG(z)-p^{-1}G(p^{-1})](1-d_1z^{-1})[z^{-1}G(z^{-1})-p^{-1}G(p^{-1})]}{(1-pz)(1-pz^{-1})}\frac{dz}{z}$$

$$+ \frac{h}{2\pi i}\oint F(z)F(z^{-1})\frac{dz}{z}$$

where $G(z) = A(z) + F(z)$. Letting $G^*(z) = (1-d_1z)[zG(z)-p^{-1}G(p^{-1})]$ we can write (B.1) as

$$J = \frac{1}{2\pi i}\oint \frac{[G^*(z)G^*(z^{-1}) + h(1-pz)(1-pz^{-1})F(z)F(z^{-1})}{(1-pz)(1-pz^{-1})z}dz.$$

Note that $\partial G^*(z)/\partial F_j = (1-d_1z)(z^{j+1} - p^{-j-1})$ and obtain the first order conditions for optimizing J by setting $\partial J/\partial F_j = 0$ for $j = 0, 1, \ldots$:

$$\begin{aligned} 0 = \frac{1}{2\pi i} \oint \Big\{ & (1-d_1 z) z^{j+1} G^*(z^{-1}) + h(1-pz)(1-pz^{-1}) z^{j} F(z^{-1}) \\ & + (1-d_1 z^{-1}) z^{-j-1} G^*(z) + h(1-pz)(1-pz^{-1}) z^{-j} F(z) \\ & - (1-d_1 z) p^{-j-1} G^*(z^{-1}) \\ & - (1-d_1 z^{-1}) p^{-j-1} G^*(z) \Big\} \Big\{ (1-pz)(1-pz^{-1}) z \Big\}^{-1} dz \end{aligned}$$

or

$$\frac{1}{2\pi i} \oint [z^{j} H(z^{-1}) + z^{-j} H(z)] \frac{dz}{z} =$$

(B.2)

$$\frac{p^{-j-1}}{2\pi i} \oint \left\{ \frac{(1-d_1 z)(1-d_1 z^{-1})[z^{-1}G(z^{-1}) - p^{-1}G(p^{-1}) + zG(z) - p^{-1}G(p^{-1})]}{(1-pz)(1-pz^{-1})} \right\} \frac{dz}{z}$$

where $H(z) = (1-d_1 z^{-1}) z^{-1} G^*(z)/(1-pz)(1-pz^{-1}) + hF(z)$. The right-hand side of (B.2) is analytic as the closed unit disk everywhere but $z = 0$. ($z = p^{-1}$ is not singularity since it factors out of the numerator). To see this, write the RHS of (B.2) as

$$\frac{p^{-j-1}}{2\pi i} \oint \left\{ \frac{(1+d_1^2-d_1 z)\ [z^{-1}G(z^{-1}) - p^{-1}G(p^{-1}) + zG(z) - p^{-1}G(p^{-1})]}{(1-pz)(1-pz^{-1})} \right\} z$$

$$- \frac{p^{-j-1}}{2\pi i} d_1 \oint \frac{zG(z) - p^{-1}G(p^{-1})}{z(1-pz)(1-pz^{-1})} \frac{dz}{z} - \frac{p^{-j-1}}{2\pi i} d_1 \oint \frac{z^{-1}G(z^{-1}) - p^{-1}G(p^{-1})}{z(1-pz)(1-pz^{-1})} \frac{dz}{z}$$

where we have expanded the $(1-d_1z)(1-d_1z^{-1})$ term and separated the terms in d_1z^{-1}. The first term in this expression is analytic everywhere inside the unit circle. The second term has a simple pole at $z = 0$, and by Cauchy's integral formula

$$
\begin{aligned}
-\frac{p^{-j-1}}{2\pi i}\, d_1 \oint \frac{zG(z)-p^{-1}G(p^{-1})}{z(1-pz)\,(z-p)}\, dz &= -p^{-j-1}d_1 \left.\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right|_{z=0} \\
&= -p^{-j-1}d_1\, \frac{p^{-1}G(p^{-1})}{p} \\
&= -p^{-j-1}d_1q_0
\end{aligned}
$$

where $q = \frac{p^{-1}G(p^{-1})}{p}$. By summetry, the third term also equals $-p^{-j-1}d_1q_0$ and as a result, $2H_j = -2p^{-j-1}d_1q_0$ for $j = 0, 1, 2, \ldots$. Multiplying by z^j and summing over $j \in [-\infty, \infty]$ results in $H(z) = \Sigma_{-\infty}^{-1} - d_1q_0p^{-1}\,\Sigma_0^{\infty}\, p^{-j}z^j$ or $H(z) = \Sigma_{-\infty}^{-1} - \frac{d_1q_0p^{-1}}{1-p^{-1}z}$. As in Whitman, $\Sigma_{-\infty}^{-1}$ is an unknown function involving only negative powers of z. Using the definition of H(x)

$$
hF(z) = \Sigma_{-\infty}^{-1} - (1-d_1z)(1-d_1z^{-1})\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] - \frac{d_1q_0p^{-1}}{1-p^{-1}z}
$$

Expanding $(1-d_1z)(1-d_1z^{-1})$ allows this to be written

$$
\begin{aligned}
hF(z) = \Sigma_{-\infty}^{-1} &- (1+d_1^2)\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] + d_1z\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] \\
&+ d_1z^{-1}\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] - \frac{d_1q_0p^{-1}}{1-p^{-1}z} .
\end{aligned}
$$

Since policy at time t may not depend on future shocks $F(z) = [F(z)]_+$ where $[\ ]_+$ is the linear annihilator which means ignore negative powers of z. As in Whiteman (1986), p. 1394) the first term on the right-hand side equals zero when the annihilator is applied, and since $z = p^{-1}$ is not a singularity of the term in brackets the second, third and fifth terms are equal to themselves. The fourth term has a simple pole at $z = 0$. By the lemma in Hansen and Sargent (1981, p. 120)

$$\left\{d_1 z^{-1}\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right]\right\}_+ = d_1 z^{-1}\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] - d_1 z^{-1} q_o$$

where $q_o = \frac{p^{-1}G(p^{-1})}{p}$. To see this note that the residue of the term in braces at $z = 0$ is given by

$$\text{(B.3)}\quad \lim_{z\to o}\left[d_1\,\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] = \frac{p^{-1}G(p^{-1})}{p}$$

and thus the principal part of the Laurent expansion of the term in braces at $z = 0$ is $d_1 p^{-1}G(p^{-1})/pz$. The Hansen-Sargent lemma requires subtracting the principal part of the Laurent expansion from the annihiland to obtain the annihilate. This is (B.3).

We can now write

$$\text{(B.4)}\quad hF(z) = -(1-d_1 z)(1-d_1 z^{-1})\left[\frac{zG(z)-p^{-1}G(p^{-1})}{(1-pz)(z-p)}\right] - d_1 z^{-1} q_o - \frac{d_1 q_0 p^{-1}}{1-p^{-1}z}$$

or

$$\text{(B.5)}\quad (1-pz)(z-p)hF(z) = -(1-d_1 z)(1-d_1 z^{-1})[zF(z) + zA(z) - p^{-1}G(p^{-1})]$$
$$- (1-pz)(1-pz^{-1})d_1 q_o - (1-pz)(1-pz^{-1})\,\frac{d_1 q_0 p^{-1}}{1-p^{-1}z}\ .$$

Notice the last term in (B.5) may be written

$$-pz(1-p^{-1}z^{-1})(-pz^{-1})(1-p^{-1}z) \cdot \frac{d_1q_0p^{-1}}{1-p^{-1}z}$$

which equals

$$p^2\,(1-p^{-1}z^{-1})d_1q_0p^{-1} = (1-p^{-1}z^{-1})d_1q_0p = d_1q_0p - z^{-1}d_1q_0 \,.$$

Thus, expanding the right-hand side and using the definition of q_o, (B.5) may be written

$$zg(1-mz)(1-mz^{-1})F(z) = -(1-d_1z)(1-d_1z^{-1})zA(z)$$
$$+ [(1+d_1^2)\,p^{-1}G(p^{-1}) - (1+p^2)d_1q_o - d_1q_0p] + z^{-1}d_1q_0$$

where

$$\text{(B.6)}\quad g(1-mz)(1-mz^{-1}) = [(1-pz)(1-pz^{-1})h + (1-d_1z)(1-d_1z^{-1})].$$

From (B.6)

$$h + hp^2 + 1 + d_1^2 = g(1+m^2) \text{ and } hp + d_1 = gm$$

which implies

$$\frac{h + hp^2 + 1 + d_1^2}{hp + d_1} = m + m^{-1}$$

which implies $|m| < 1$. Thus, we can write (B.5) as

$$\text{(B.7)}\quad g(1-mz)F(z) = -(1-mz^{-1})^{-1}D(z)$$
$$+ (1-mz^{-1})^{-1}z^{-1}[(1+d_1^2)p^{-1}G(p^{-1})-(1+p^2)d_1q_o - d_1q_0p$$
$$+ (1-mz^{-1})^{-1}\,z^{-2}d_1q_0]_+$$

where $D(z) = (1-d_1z)(1-d_1z^{-1})A(z)$. Now apply the linear annihilator operator $[\ ]_+$ to both sides of (B.7) noting that the left hand side function contains no negative powers of z. So

$$g(1-mz)F(z) = -[(1-mz^{-1})^{-1}D(z)]_+ +[(1-mz^{-1})^{-1}z^{-1}((1+d_1^2)p^{-1}G(p^{-1})-(1+p^2)d_1q_o$$
$$-d_1q_0p) + (1-mz^{-1})^{-1}\,z^{-2}d_1q_0]_+$$

The last term vanishes as it involves only negative powers of z. Therefore the z-transform of F(L) is given by

$$F(z) = -g^{-1}(1-mz)^{-1}[(1-mz^{-1})^{-1}D(z)]_{+}.$$

Again, by the lemma in Hansen and Sargent (1980, p. 120)

$$[(1-mz^{-1})^{-1}D(z)]_{+} = -m^{-1}\frac{zD(z) - mD(m)}{1 - m^{-1}z}.$$

Noting that $gm = (hp + d_1)$ completes the proof.

Appendix C

DATA**

The data used for estimating the Maine potato model is given in the following table. The soil management measure is given by the formula SM=1/(1-PO) where PO is the percentage of total acreage of oats and potatoes planted in oats. This gives a measure of soil management on a scale from 0 to infinity. The suggested rotation is oats, potatoes, yearly, thus the "goal" value of SM is 2 (PO=0.5). It is a good measure for soil management because the P factor in the USLE has remained essentially constant over the period of the data (see Hepler, et al, 1985). Over the data period, the primary force determining the C factor has been crop rotation, and thus SM should be closely correlated with overall soil management.

Nonsoil inputs (INPUTS) is a composite measure of purchased and nonpurchased farm inputs for the Northeast. Soil depth (SOILD) was estimated using the USLE. See Lawrence for details. The relative price of inputs (INPRI) is the ratio of an index of prices paid by farmers in the Northeast and potato prices for the state of Maine. Finally, the government

---

** I would like to thank Doug Lawrence for providing the data used in this analysis. For a complete description see Lawrence (1987).

subsidy, ACP, is the ratio of ACP cost sharing per acre for the state of Maine to the price of potatoes. All nominal dollars were adjusted to 1967 values using the Consumer Price Index.

ENTRY	INPUTS	SM	SOILD	ACP	INPRI
1950:	146.000	1.70455	11.5700	.503015	190.698
1951:	145.000	2.31000	11.5540	.207164	90.099
1952:	143.000	1.66667	11.5460	.280420	126.267
1953:	141.000	1.71429	11.5300	.756113	345.946
1954:	138.000	1.75000	11.5150	.254707	118.605
1955:	135.000	1.69504	11.5000	.319126	141.808
1956:	133.000	1.61905	11.4840	.461838	206.612
1957:	128.000	1.75912	11.4670	.247033	119.535
1958:	126.000	1.57823	11.4530	.489282	237.838
1959:	124.000	1.63121	11.4340	.232384	114.655
1960:	121.000	1.48299	11.4170	.408122	194.853
1961:	119.000	1.38095	11.3860	.518767	235.398
1962:	117.000	1.41781	11.3570	.463293	221.311
1963:	115.000	1.42336	11.3310	.305567	141.451
1964:	112.000	1.38571	11.3050	.149953	70.130
1965:	110.000	1.35099	11.2760	.228481	117.373
1966:	108.000	1.26582	11.2430	.317647	170.000
1967:	107.000	1.23602	11.2050	.395359	213.235
1968:	104.000	1.25161	11.1580	.289774	156.757
1969:	102.000	1.25466	11.1150	.229523	137.273
1970:	102.000	1.30719	11.0720	.257576	158.081
1971:	100.000	1.31034	11.0350	.309598	192.941
1972:	97.0000	1.34074	11.0000	.130710	85.610
1973:	98.0000	1.31884	10.9680	.074270	58.483
1974:	99.0000	1.31690	10.9330	.182556	165.862
1975:	97.0000	1.40164	10.8980	.093760	87.273
1976:	99.0000	1.32759	10.8700	.129870	112.929
1977:	100.000	1.30645	10.8360	.183715	172.321
1978:	103.000	1.38655	10.8000	.161721	162.694
1979:	106.000	1.38793	10.7710	.202580	221.538
1980:	104.000	1.42593	10.7420	.093148	110.207
1981:	102.000	1.46226	10.7160	.146237	190.000
1982:	101.000	1.43925	10.6920	.195774	258.209
1983:	97.0000	1.43750	10.6670	.119135	149.831

References

S. Batie, Policies, institutions and incentives for soil conservation, in "Soil Conservation Policies, Institutions and Incentives," (Halcrow, et al, eds.), Soil Conservation Society of America, Ankeny, Iowa, (1982).

G. E. P. Box and G. M. Jenkins, "Time Series Analysis, Forecasting and Control," Holden Day, San Francisco (1970).

D. W. Bromley, The rights of society versus the rights of landowners and operators, in "Soil Conservation Policies, Institutions and Incentives,"(Halcrow, et al, eds.), Soil Conservation Society of America, Ankeny, Iowa, (1982).

O. R. Burt, Farm level economics of soil conservation in the Palouse area of the Northwest, American Journal of Agricultural Economics. 63, 83-91 (1981).

K. W. Easter and M. L. Cotner, Evaluation of current soil conservation strategies, in "Soil Conservation Policies, Institutions and Incentives," (H. G. Halcrow, et al, eds.), Soil Conservation Society of America, Ankeny, Iowa (1982).

B. Eleveld and H. G. Halcrow, How much soil conservation is optimal for society?," in "Soil Conservation Policies, Institutions and Incentives," (H. G. Halcrow, et al, eds.), Soil Conservation Society of Amaerica, Ankeny, Iowa, (1982).

C. Ervin and D. Ervin, Factors affecting the use of soil conservation practices: hypotheses, evidence and policy implications, Land Economics, 58, 277-292, (1982).

D. L Forster and G. S. Becker, Costs and income effects of alternative erosion control strategies: the Honey Creek watershed, North Central Journal of Agricultural Economics, 1, 53-60, (1979).

J. E. Freund, "Mathematical Statistics," 2nd ed., Prentice-Hall, Englewood Cliffs, N.J. (1971).

R. E. Heimlich and N. Bills, An improved erosion classification for conservation policy, Journal of Soil and Water Conservation, July, 261-266, (1984).

P. R. Hepler, L. H. Long and J. A. Ferwerda, "FARMS" crop yield and quality relationships with soil erosion, 1980., bulletin 799, Maine Agricultural Experiment Station, University of Maine, Orono, Maine (1983).

G. G. Judge, et al, "The Theory and Practice of Econometrics," 2nd ed., Wiley, New York (1985).

D. J. Lawrence, "Resource Use Rates: Estimation of Time Invariant Decision Rules," unpublished PhD dissertation, University of New Hampshire, (1987).

R. E. Lucas, Jr., Econometric Policy Analysis: a critique, in "The Phillips Curve and Labor Markets," (K. Brunner and S. H. Meltzer, eds.), North-Holland, Amsterdam, (1976).

K. E. McConnell, An Economics Model of Soil Conservation, American Journal of Agricultural Economics, 65, 83-89, (1983).

I. C. Moore, et al, Financial incentives to control agricultural nonpoint-source pollution, Journal of Soil and Water Conservation, March, 60-64, (1979).

T. L. Napier and D. L. Forster, Farmer attitude and behavior associated with soil erosion control, in "Soil Conservation Policies Institutions and Incentives,' (Halcrow, et al eds.), Soil Conservation Society of America, Ankeny, Iowa, (1982).

T. L. Napier, et al, Factors affecting adoption of conventional and conservation tillage practices in Ohio, Journal of Soil and Water Conservation, May, 205-209, (1984).

T. Prato, Allocation of federal assistance to soil conservation, Land Economics, 63:2, 193-200, (1987).

W. D. Rasmussen, History of soil conservation, institutions and incentives, in "Soil Conservation Policies, Institutions and Incentives," (H. G. Halcorw, et al, eds.), Soil Conservation Society of America, Ankeny, Iowa, (1982).

R. E. Rosenman, Soil management, productivity and expectations, mimeo, Washington State University, (1986).

B. Saliba and D. Bromley, Empirical analysis of the relationship between soil conservation and farmland characteristics, selected paper, AAEA meetings, Ithaca, New York, (1984).

J. S. Shortle and J. A. Miranowski, Intertemporal soil resource use: is it socially excessive?, Journal of Environmental Economics and Management,14, 99-111, (1987).

C. H. Whiteman, "Linear Rational Expectations Models: A Users Guide," University of Minnesota Press, Minneapolis, (1983).

C. H. Whiteman, Analytical policy design under rational expectations, Econometrica, 6, 1387-1405, (1986).

W. H. Wischmeier and D. D. Smith, "Predicting Rainfall Erosion Losses: A Guide to Conservation Planning," USDA Agricultural Handbook No. 537, (1978).

Table 1: INPUTS

```
DEPENDENT VARIABLE   13      INPUTS
FROM  1952:  1  UNTIL  1983:  1
OBSERVATIONS              32     DEGREES OF FREEDOM     28
R**2               .92101723     RBAR**2            .91255479
SSR          506.71397           SEE         4.2540484
DURBIN-WATSON 1.40246133
Q(  15)=    9.00915       SIGNIFICANCE LEVEL  .877039
LABEL      VAR  LAG   COEFFICIENT    STAND. ERROR    T-STATISTIC
*****      ***  ***   ************   ************    ************
CONSTANT     0    0   -235.1898       33.06478        -7.113000
SM          14    1    35.33967       4.366006         8.094280
SOILD        4    0    26.11780       3.286415         7.947202
INPRI       15    1    .2719733E-01   .1235583E-01     2.201174
```

Table 2A: SOIL MANAGEMENT

```
DEPENDENT VARIABLE   14      SM
FROM  1952:  1  UNTIL  1982:  1
OBSERVATIONS              31     DEGREES OF FREEDOM     27
R**2               .82716479     RBAR**2            .80796087
SSR          .12934168           SEE         .69212951E-01
DURBIN-WATSON .87587781
Q(  15)=    23.3652       SIGNIFICANCE LEVEL  .766883E-01
LABEL      VAR  LAG   COEFFICIENT    STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************    ************
CONSTANT     0    0   3.219786        .9043187         3.560455
SOILD        4   -1   -.3149353       .9673736E-01    -3.255571
INPUTS      13   -1   .1515624E-01    .1841353E-02     8.231038
ACP         16    1   .1498358        .1091318         1.372980
```

Table 2B: SOIL MANAGEMENT (CORRECTED FOR SERIAL CORRELATION)

```
DEPENDENT VARIABLE   14      SM
FROM  1953:  1  UNTIL  1982:  1
OBSERVATIONS              30     DEGREES OF FREEDOM     26
R**2               .87709435     RBAR**2            .86291293
SSR          .85368105E-01       SEE         .57300861E-01
DURBIN-WATSON 1.74190968
Q(  15)=    9.51948       SIGNIFICANCE LEVEL  .848829
LABEL      VAR  LAG   COEFFICIENT    STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************    ************
CONSTANT     0    0   3.325983        1.328401         2.503750
SOILD        4   -1   -.3311274       .1453958        -2.277420
INPUTS      13   -1   .1585510E-01    .3394528E-02     4.670782
ACP         16    1   .1765368        .8125106E-01     2.172733

LABEL      VAR  LAG   COEFFICIENT    STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************    ************
RHO          1    0   .5348958        .1788211         2.991235
```

Table 3A: SOIL DEPTH

```
DEPENDENT VARIABLE     4      SOILD
FROM  1952:  1  UNTIL  1983:  1
OBSERVATIONS                 32     DEGREES OF FREEDOM     30
R**2              .99973054     RBAR**2            .99972156
SSR         .70685309E-03     SEE           .48540467E-02
DURBIN-WATSON .96092942
Q(  15)=    31.4270         SIGNIFICANCE LEVEL  .769810E-02
LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
SOILD       4    1   .9917293         .5803566E-03   1708.828
SM         14    1   .4417327E-01    .4379082E-02   10.08734
```

Table 3B: SOIL DEPTH (CORRECTED FOR SERIAL CORRELATION)

```
DEPENDENT VARIABLE     4      SOILD
FROM  1953:  1  UNTIL  1983:  1
OBSERVATIONS                 31     DEGREES OF FREEDOM     29
R**2              .99990025     RBAR**2            .99989681
SSR         .24531739E-03     SEE           .29084739E-02
DURBIN-WATSON 1.24954366
Q(  15)=    13.3688         SIGNIFICANCE LEVEL  .573832
LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
SOILD       4    1   .9891213         .5201961E-03   1901.439
SM         14    1   .6475179E-01    .4027611E-02   16.07697

LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
RHO         1    0   .3513353E-01    .8421299E-01   .4171985
```

Table 4: ACP (ACTUAL SUBSIDY RULE)

```
DEPENDENT VARIABLE   16     ACP
FROM  1952:  1  UNTIL  1983:  1
OBSERVATIONS             32     DEGREES OF FREEDOM     21
R**2             .96456216     RBAR**2          .94768699
SSR          .24734960E-01     SEE          .34319896E-01
DURBIN-WATSON 1.96013797
Q(  15)=    10.5517        SIGNIFICANCE LEVEL  .783682
LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
CONSTANT    0    0  -5.072464        1.313992       -3.860347
ACP        16    1  -.3611584        .2268784       -1.591859
ACP        16    2  -.2394822E-01    .5874922E-01   -.4076347
INPUTS     13    0   .1955010E-03    .4669734E-02    .4186556E-01
INPUTS     13    1   .8150669E-03    .5308226E-02    .1535479
SM         14    0  -.8802489E-01    .1443231       -.6099154
SM         14    1  -.1019312        .8994309E-01   -1.133285
SOILD       4    0  -.3932119        1.895465       -.2074488
SOILD       4    1   .8618584        1.905858        .4522156
INPRI      15    0   .1650003E-02    .1406665E-03   11.72989
INPRI      15    1   .5775978E-03    .4290431E-03    1.346247
```

Table 5A: ACP (CONSISTENT RULE)

```
DEPENDENT VARIABLE   16      ACP
FROM  1952:  1  UNTIL  1983:  1
OBSERVATIONS              32      DEGREES OF FREEDOM      29
R**2            -.20386554      RBAR**2          -.28689075
SSR           .84027592         SEE           .17022054
DURBIN-WATSON 1.49400504
Q(  15)=    30.5949        SIGNIFICANCE LEVEL  .994831E-02
LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
CONSTANT    0    0  -.9885730        .4736505       -2.087136
ACP        16    1  -.3280431E-01    .2590371       -.1266395
SM         14    0   .8808595        .3558537        2.475341
```

Table 5B: ACP (CONSISTENT RULE ADJUSTED FOR SERIAL CORRELATION)

```
DEPENDENT VARIABLE   16      ACP
FROM  1953:  1  UNTIL  1983:  1
OBSERVATIONS              31      DEGREES OF FREEDOM      28
R**2             .20761079      RBAR**2           .15101156
SSR           .55297775         SEE           .14053187
DURBIN-WATSON 1.72396395
Q(  15)=    10.2734        SIGNIFICANCE LEVEL  .802196
LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
CONSTANT    0    0  -.1536950        .1936818       -.7935442
ACP        16    1   .5062937        .2237805        2.262457
SM         14    0   .1990600        .1529790        1.301224

LABEL      VAR  LAG    COEFFICIENT   STAND. ERROR    T-STATISTIC
*******    ***  ***   ************   ************   ************
RHO         1    0  -.3310390        .2403910       -1.377086
```

Table 6A: CHI-SQUARE TEST OF INPRI AS WHITE NOISE PROCESS

```
DEPENDENT VARIABLE   15      INPRI
FROM  1952:  1  UNTIL  1983:  1
OBSERVATIONS              32      DEGREES OF FREEDOM     31
R**2              .00000000      RBAR**2           .00000000
SSR          115406.48           SEE          61.014667
DURBIN-WATSON  2.13266223
Q(  15)=    7.77015        SIGNIFICANCE LEVEL  .932666
LABEL      VAR  LAG     COEFFICIENT    STAND. ERROR    T-STATISTIC
*******    ***  ***     ************   ************    ************
CONSTANT    0    0      162.3445       10.78597        15.05145
```

Table 6B: AUTOCORRELATIONS FOR INPRI (RELATIVE INPUT PRICE VARIABLE)

```
    34 Observations, From 1950 to 1983

              LAG      ESTIMATE      z VALUE
               1       -.065129       -.36
               2       -.180077       -.97
               3        .145610        .74
               4       -.087320       -.49
               5        .160926        .81
               6       -.208028      -1.17
```

Table 7: SIMULATION WITH ACTUAL RULE

ENTRY	SOIL MGMT	SOIL DPTH	ACP
8	1.59219	11.4589	.284177
9	1.63582	11.4373	.476799
10	1.58828	11.4188	.262810
11	1.57601	11.3974	.392827
12	1.66572	11.3755	.444937
13	1.66991	11.3596	.398986
14	1.56801	11.3442	.273326
15	1.44052	11.3223	.170683
16	1.43946	11.2924	.250670
17	1.51909	11.2627	.315329
18	1.58411	11.2386	.360088
19	1.54211	11.2189	.256743
20	1.47559	11.1967	.227283
21	1.47898	11.1704	.259528
22	1.52195	11.1447	.301103
23	1.42719	11.1220	.116825
24	1.31839	11.0934	.842970E-01
25	1.40527	11.0581	.255672
26	1.39140	11.0288	.100784
27	1.34724	10.9989	.140116
28	1.41632	10.9665	.229675
29	1.43804	10.9389	.190166
30	1.47541	10.9130	.277749

Table 8: SIMULATION WITH OPTIMAL RULE

ENTRY	SOIL MGMT	SOIL DPTH	ACP
7	2.86322	11.6141	.754200
8	2.20055	11.6994	.434701
9	2.99948	11.7409	.835244
10	1.97217	11.8337	.344028
11	2.76963	11.8590	.740128
12	2.56516	11.9356	.639075
13	2.30575	11.9982	.507979
14	2.00804	12.0433	.362712
15	2.01840	12.0686	.356068
16	2.47367	12.0943	.580694
17	2.46050	12.1492	.570345
18	2.38845	12.2027	.533766
19	1.95337	12.2509	.320581
20	2.08211	12.2704	.383581
21	2.22128	12.2980	.452641
22	2.25128	12.3344	.464268
23	1.64900	12.3723	.165114
24	1.96104	12.3708	.310359
25	2.48522	12.3895	.568315
26	1.70089	12.4419	.186482
27	2.10974	12.4430	.382252
28	2.22366	12.4705	.436364
29	1.91962	12.5052	.293423
30	2.17939	12.5197	.421077

Table 9: YEARLY VALUE OF THE REGULATORY OBJECTIVE FUNCTION (SIMULATION)

ENTRY	OPTIMAL RULE	ACTUAL RULE
8	.251068	.226517
9	1.33035	.209264
10	.499167	.264097
11	.916393	.214680
12	.340501	.117353
13	.128979	.113324
14	.043652	.219232
15	.042961	.334782
16	.328585	.327417
17	.212282	.239912
18	.153661	.177106
19	.096046	.231722
20	.014939	.276800
21	.058814	.273608
22	.063421	.232102
23	.308045	.398256
24	.045092	.466774
25	.372879	.414364
26	.390611	.419947
27	.091205	.429297
28	.056073	.357253
29	.048663	.319021
30	.065841	.291044
mean	.252899	.284950

FIGURE 1: SHOCK TO SOIL MANAGEMENT

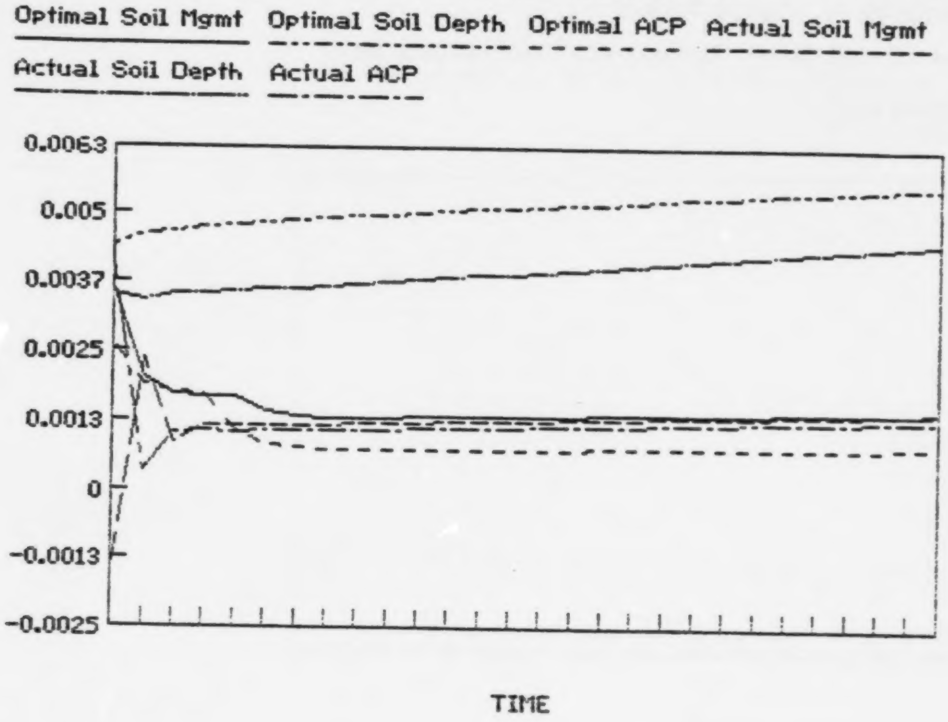


FIGURE 2: SHOCK TO SOIL DEPTH

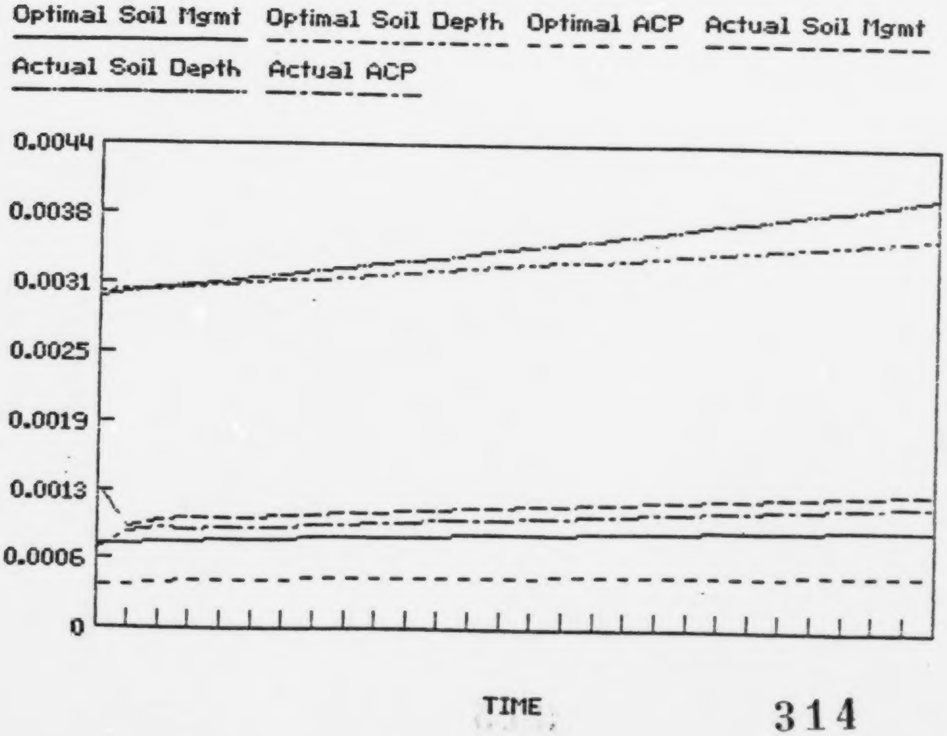


FIGURE 3: SHOCK TO ACP

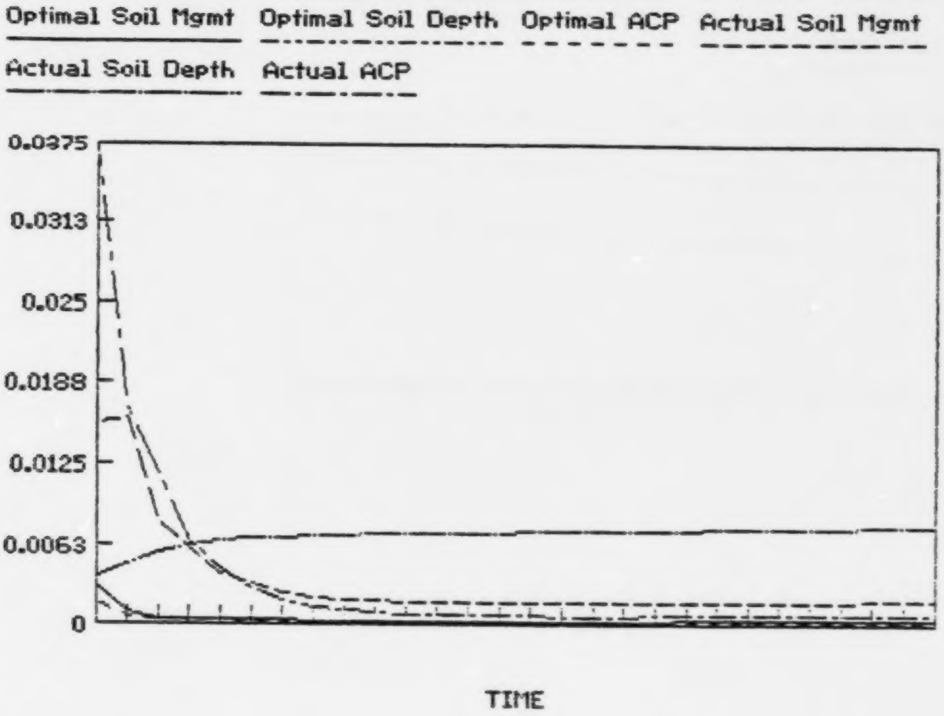


FIGURE 4: SOIL MANAGEMENT
SIMULATED VALUES

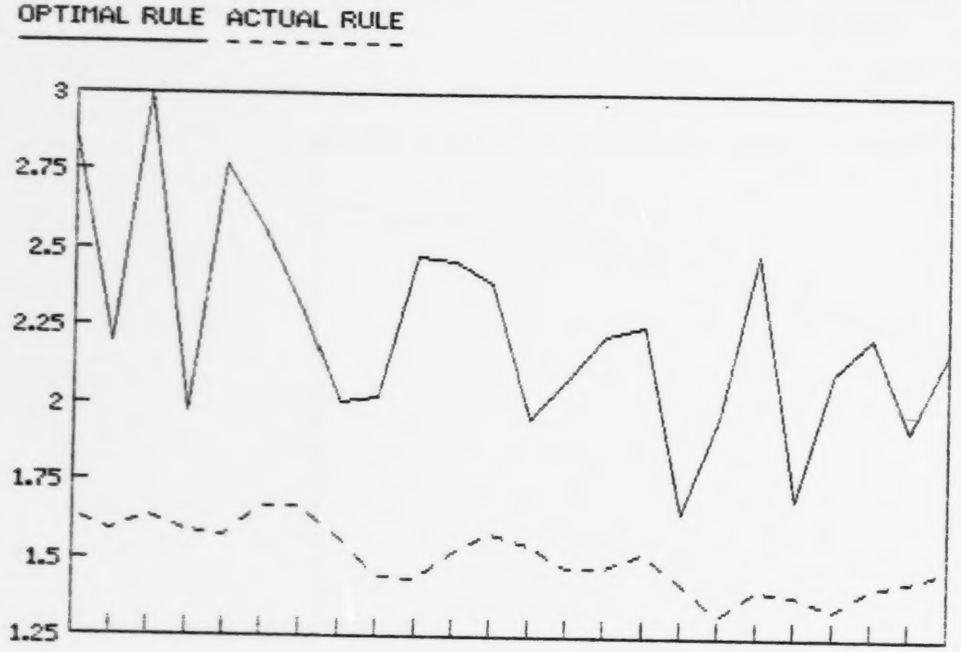


FIGURE 5: ACP
SIMULATED VALUES

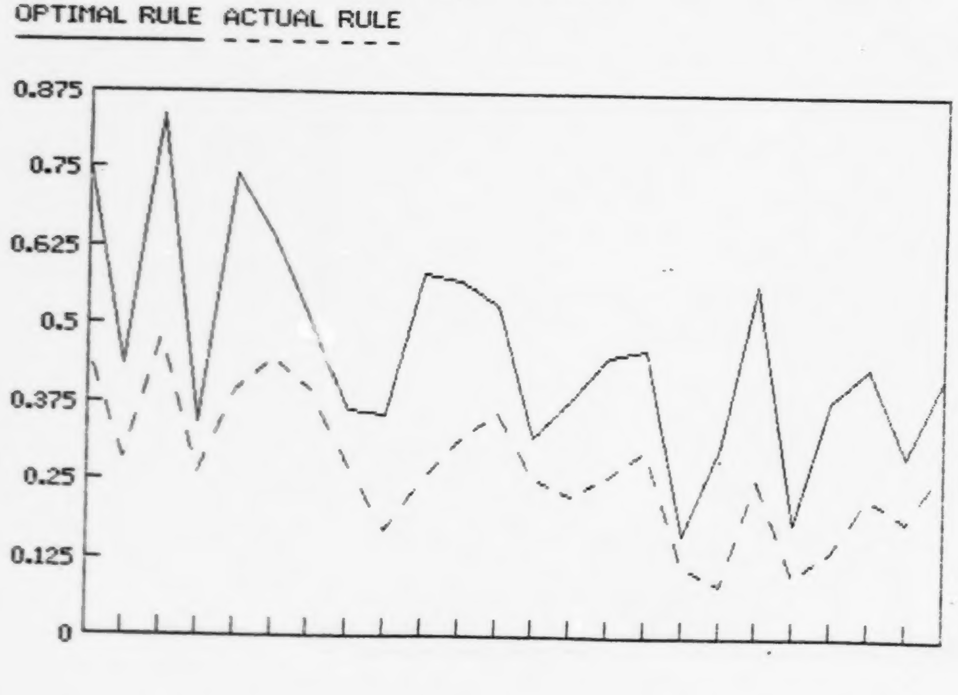
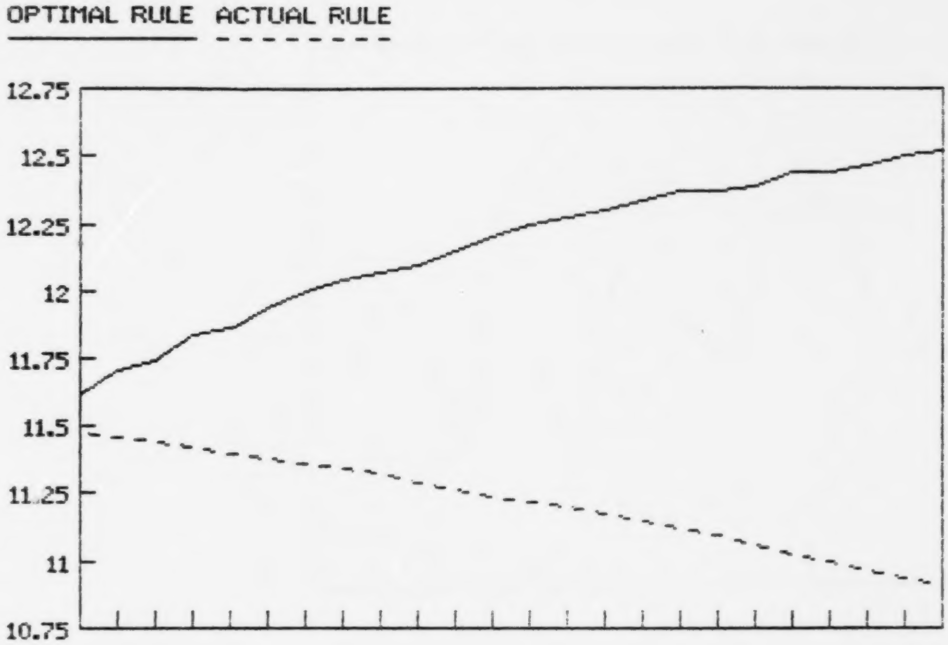
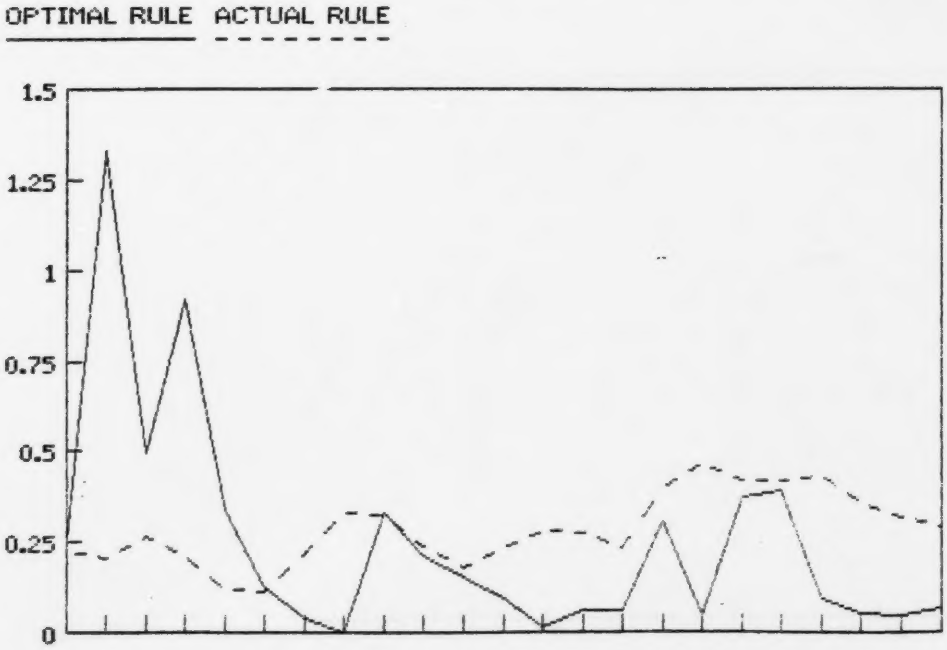


FIGURE 6: SOIL DEPTH
SIMULATED VALUES



@

FIGURE 7: OBJECTIVE FUNCTION VALUE
SIMULATED VALUES



@

1. The ACP is administered through the Agricultural Stabilization and Conservation Service (ASCS). The SCS provides mostly technical assistance, while the ACP is a program of subsidies and cost-sharing.

2. There is a large literature discussing optimal resource use rates. Specific applications to soil erosion are more limited. Burt (1981) and McConnell (1983) use dynamic models of farmer decisions to study soil erosion and its relationship to society's preferred level. Bromley (1982) discusses the difference between what society wants and what farmers find to be optimum. And Eleveld and Halcrow (1982) do an analysis based on Pareto optimality conditions. In a more recent analysis, Shortle and Miranowski (1987) focus on the offsight impacts of soil erosion, and the social loss that accompanies it.

3. See Whiteman for a longer discussion of the difference between a time consistent rule and a precommitment one. Much of the application in this paper was inspired by his derivations in a generic model of optimal policy formulation. (Whiteman terms the time consistent policy the "mercurial" policy).

4. There are too many studies of this sort to cite all of them. Primary examples include Ervin and Ervin (1982), Napier and Forster (1982), and Saliba and Bromley (1984).

5. This description is based on the one in Moore, et al, (1979).

6. See Wischmeier and Smith (1978) for a discussion of soil tolerance levels. For our purposes it is necessary to note that soil erosion is of two primary types, rill and sheet, and is effected by the physical constraints (rainfall, soil type, field slope and shape) and management practices (essentially crop rotation and cultivation).

7. See Wischmeier and Smith (1978) or Heimlich and Bills (1984) for discussions of the USLE.

8. It is important to differentiate between T, the tons lost per acre, and "T", the tolerance level of erosion. The former is actual erosion, the latter is the normative value that soil scientists believe the soil can tolerate without suffering long-term productivity effects.

9. For notational convenience the function f(.,.,.) is unspecified here. In the first appendix a quadratic approximation is used to derive the expectational difference equation (7) which characterizes the farmer's choice of soil management through time. Rosenman (1986) includes an

estimation of a quadratic production function for different data.

10. This is actually the optimal precommitment rule if $M*=S*=0$. In the application the rule is derived for demeaned values, thus adhering to this condition, and then adjusted to level.

11. If $h=0$, then $d(m)d(m^{-1})=0$ and $m+m^{-1}=(1+d_1^2)/d_1$. Thus

$$F(z) = \{[zd(z)d(z^{-1})]/[d_1(1-mz)(1-m^{-1}z)]\}A(z).$$

Expanding the coefficient on $A(z)$ shows it equals -1.

12. The derivation of equations (12) and (13) are given in Appendix A.

13. Notice that if $d_1=0$ equation (13) reduces to Whiteman's closed loop precommitment policy.

14. All of the calculations in this paper were carried out using the VAR Econometrics, Inc. software RATS. The equations were estimated by two-stage-least-squares using lagged soil depth, input prices and government subsidies. The path of relative prices was analyzed using OLS.

15. A description of the data is provided in Appendix C.

16. Let the optimization problem, equation (2) be

$$J = E\{M_t-M*)^2 + h(S_t-S_{t-1})\}.$$

Then the first order condition to maximize J is its derivative with respect to S_t set equal to zero, taking as given equation

(7):

$$2h(S_t - S_{t-1}) + 2(M_t - M*)(K_3/[d_1 + K_1 d_2]) = 0$$

where the term in the third set of parentheses is derived using the implicit function theorem. Rearranging gives the equation estimated in Table 5A.

17. The negative R^2 comes from an algorithm that computes this value using GLS estimates of the errors. It is also indicative of serial correlation in the errors. See Judge, et al, (1985, p.30 and p.477).

18. The chi-square test was performed by regressing INPRI on a constant, and using the reported Box-Pierce Q-statistic to check for serial correlation. The reported statistic was 7.77, indicating no serial correlation.

19. Under the assumption that farmers are competitors in both the inputs and output markets, which allowed the specification with yield as a numeraire, this is not a surprising conclusion. If price shocks in the input and output markets are (uncorrelated) white noises, then their ratio should also follow a white noise around some central value. This is argument is strengthened when one realizes that Maine potatoes as a whole, the market used for these estimations, are often thought to be price takers against the potatoes grown in western markets.

20. It is important to realize that a negative shock on ACP in the actual system would have a strong, persistent and growing negative affect on soil depth, thus in times of budgetary crisis could have very deleterious outcomes on soil erosion.

The Effectiveness of EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards

by

Wesley A. Magat
Fuqua School of Business
Duke University

and

W. Kip Viscusi
Department of Economics
Northwestern University

May, 1987

*This work was completed under two Cooperative Agreements from the U.S. Environmental Protection Agency, one to Duke University (#CR811902-02) and one to Northwestern University (#CR81302-01). The authors thank Alan Carlin, who was invaluable in facilitating the research process, and the many people at EPA who so kindly assisted us with obtaining the data and understanding the agency's enforcement process. Anil Gaba provided superb computer programming support.

Abstract

This study evaluates the effect of EPA water pollution standard enforcement on pollution in the pulp and paper industry. The original data set used consisted of a longitudinal data base of 77 pollution sources tracked on a quarterly basis from 1982:1 to 1985:1. EPA inspections have a statistically significant effect in the expected direction on pollution discharge levels, the probability of non-compliance, and the regularity of filing discharge monitoring reports. Although these effects are consistent with EPA's legislative mandate, they probably pass a benefit-cost test only for firms that make no capital expenditures and only modest operating expenditures following an inspection.

I. Introduction

In the almost two decades since the initial wave of social regulation, the academic literature has documented very few if any instances of a health, safety, or environmental regulation that have been an unqualified success. Indeed, in most cases the problem is even more fundamental. The typical analysis of government regulation has found that the regulation did not even fulfill its primary mission, much less pass some kind of more demanding benefit-cost test.

This absence of a well-documented case study of effective social regulation may be due in part to the particular set of regulations that has been selected for analysis. There is certainly no inherent economic reason why such regulations cannot play a productive role in our economy. In the case of environmental quality, for example, the externality problems being addressed are not handled well by markets, implying that government regulation has at least the potential for playing a beneficial role. However, this potential will fail to be realized if the regulations are ill-conceived, are not effectively enforced, or if the environmental problem has no feasible solution.

A brief review of past regulatory experiences may be instructive to put the Environmental Protection Agency's (EPA) water pollution control effort, which is the focus of this paper, in better perspective. It should be noted that most of these detailed evaluations have been done with respect to agencies other than EPA. Although there have been some treatments of EPA regulations in the academic literature,[1] as well as some assessments within the government,[2] none of these evaluations have been undertaken with the same degree of statistical rigor and detailed empirical analysis that has characterized analyses of health and safety regulations.

In large part this lack of attention stems from the greater difficulty in constructing an environmental data base.[3] The decentralized nature of polluting activity, some of which is clandestine, makes pollution levels more difficult to monitor than compliance with, for example, safety cap requirements. These difficulties posed for external evaluation may also generate monitoring problems for the agency's enforcement staff. An important issue to be addressed here is whether the prolonged process required for us to amass a sound environmental data base for the purpose of external analysis is a reflection of underlying intrinsic difficulties in the monitoring and enforcement of EPA regulations.

The past assessments of health and safety regulations have indicated that regulations have been ineffective in promoting their objectives for two general classes of reasons. The first of these is ineffectively designed regulatory policies. Thus, even though there is compliance with the regulatory requirements, little or no beneficial effect has been observed.

---

[1] Robert W. Crandall, Controlling Industrial Pollution: The Economics and Politics of Clean Air (1983); Paul MacAvoy, The Regulation of Air Pollutant Emissions from Plants and Factories (1981); and B. Peter Pashigian, Environmental Regulation: Whose Self-Interests are being Protected?, 23 Econ. Inquiry 551 (1985), are excellent examples of such contributions.

[2] See, for example, U.S. General Accounting Office, Wastewater Dischargers Are Not Complying with EPA Pollution Control Permits (1983).

[3] See Robert W. Crandall, Controlling Industrial Pollution: The Economics and Politics of Clean Air (1983), for discussion of many of the problems confronted with respect to air pollution data.

The seat belt requirements of the National Highway Traffic Safety Administration are one exhaustively studied instance of this type. Because many drivers do not use seat belts, and those that do may alter driving habits, the regulation has not produced the dramatic effect that the proponents of the regulation envisioned. Although some studies suggest that there has been no significant effect,[4] while others suggest a modest beneficial effect,[5] the overall implication is that there has not been a major effect of seat belts because the crucial behavioral link involving drivers was not considered by those designing the policy.

A similar effect has been observed with respect to the Consumer Product Safety Commission's safety requirements,[6] and more generally there is evidence that consumer product safety regulations are not sufficiently effective or extensive to have a substantial effect on product safety. Manufacturers have complied with the regulatory standards, but consumer safety has not been enhanced.

---

[4] For supporting data, see Sam Peltzman, The Effects of Auto Safety Regulation, 83 J. Pol. Econ. 677 (1975).

[5] Among the best of the optimistic assessments of seat belt regulations is that of Robert W. Crandall and John D. Graham, Automobile Safety Regulation and Offsetting Behavior: Some Empirical Estimates, 74 Amer. Econ. Rev. 328 (1984).

[6] See W. Kip Viscusi, Consumer Behavior and the Safety Effects of Product Safety Regulation, 18 J. Law & Econ. 527 (1985).

Much the same story is true in the pharmaceutical area. Pharmacists and doctors have complied with the U.S. prescription requirements for drugs, with only occasional notable violations. Nevertheless, in terms of the effect of prescriptions on health, no significant health effects of these requirements have been observed either for the United States or elsewhere in the world.[7]

The second reason for regulatory failure is the lack of enforcement. For example, the Occupational Safety and Health Administration (OSHA), has extensive regulatory requirements but traditionally has enforced them quite laxly. Indeed, the inspection rates are so low (less than one inspection per century per firm) and the penalties are so small (only $6 million annually) that there are few incentives for compliance. The result is that there has been at best a very modest effect on safety outcomes.[8]

The EPA water pollution regulations, which will be the focus of this study, represent an interesting departure from past patterns of regulatory failure. First, the nature of the regulations -- discharge limits -- is directly related to the policy objective of controlling pollution, and there is no potential for offsetting behavioral responses. If the pollution standards are binding and if they are enforced, they should improve water quality. Second, the enforcement effort is so extensive that there should be an effect of enforcement on firms' compliance. In the pulp and paper industry, which we will analyze, EPA averages roughly one inspection annually. In addition, firms are required to file monthly discharge monitoring reports, providing one of the most thorough monitoring capabilities of any health, safety, or environmental agency. One of the potential weak links is that EPA officials cannot assess penalties for non-compliance directly. They can, however, seek the imposition of substantial penalties through court action.

---

[7] For supporting data, see Sam Peltzman, The Health Effects of Mandatory Prescriptions, J. Law & Econ. (forthcoming).

[8] By the most extensive analysis is that in W. Kip Viscusi, The Impact of Occupational Safety and Health Regulation, 1973-1983, 17 Rand J. Econ. 567 (1986). Analysis of earlier periods of OSHA enforcement is provided in Ann Bartel and Lacy Thomas, Direct and Indirect Effects of OSHA Regulations, 28 J. Law & Econ. 1 (1985), and in Robert S. Smith The Impact of OSHA Inspections on Manufacturing Injury Rates, 14 J. Human Resources 145 (1979).

In the subsequent sections, we will describe the nature of the EPA enforcement of water pollution regulations in the pulp and paper industry and the original data base we created for this study. Using information from EPA and industry sources, we constructed a longitudinal data base by firm that permits a detailed evaluation of the effects of EPA inspections, and their associated enforcement actions, on the behavior of pulp and paper plants. As the empirical results will indicate, we find diverse evidence of significant EPA effects on the polluting and reporting activities of firms in the pulp and paper industry.

II. Enforcement of Water Pollution Regulations in the Pulp and Paper Industry

In choosing to study the enforcement of environmental regulations by the U. S. Environmental Protection Agency and by state environmental agencies, we could have chosen several different media. Only for water pollution was it possible to find a relatively complete data base of pollution discharge measurements by source and a data base on enforcement actions at these same plants. The same informational base that permits us to provide a sound empirical analysis also assists EPA in its effort to monitor and enforce compliance. Overall, it is believed that more than 90 percent of all major water discharges are in compliance with EPA standards, as contrasted with estimated compliance rates as low as 20 percent for toxic and hazardous substance regulation.[9] Thus, one should be cautious in generalizing the record of EPA in the water pollution area to other types of pollution problems. The investigation reported here should be regarded as an examination of an important and representative component of one of EPA's most effective regulatory programs.

Since the data on inspections were much more complete than on other enforcement actions, such as administrative orders, notices of violations, warning letters, and telephone calls,[10] we focus on the relationship between plant inspections and water pollution discharge levels. This emphasis on inspections also accords with our a priori views regarding the role of different enforcement instruments since inspections are one of the most important components of any enforcement program and thus merit special attention.

---

[9] For supporting data, see Cheryl Wasserman, Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies, United States: A National Review, O.E.C.D. (1985).

To measure the relationship between inspections and subsequent compliance, we examine one industry, pulp and paper. This industry is the country's largest discharger of conventional pollutants, such as organic waste and sediment,[11] and it has a long history of water pollution enforcement efforts by various governmental agencies. There is no reason to believe that the effectiveness of inspections in the pulp and paper industry differs markedly from that in other industries regulated by EPA. Also, by concentrating on one industry, we avoid the problem of controlling for inter-industry differences in the stringency of regulations, differences in the nature of the pollution, and differences in the technologies for compliance.

EPA has traditionally focused on the control of Biological Oxygen Demand (BOD) because it is the most damaging conventional pollutant discharged by the pulp and paper industry.[12] Most inspections examine BOD levels in addition to other pollutants of interest for a given plant. Also, the technologies which control BOD discharges tend to reduce the levels of other pollutants, which means that the relationship between inspections and BOD discharge reductions ought to be similar to the relationship between inspections and discharge reductions for other pollutants.

---

[10] One reason for the completeness of the data on inspections is that the EPA regional offices are not credited with conducting an inspection until it is coded into the central data base. See page iii in U.S. Environmental Protection Agency, Office of Water Enforcement, NPDES Inspection Manual (June, 1984).

[11] U.S. General Accounting Office, Water Pollution: Application of National Cleanup Standards to the Pulp and Paper Industry (March, 1987), page 8.

The pulp and paper industry consists of hundreds of companies operating plants in 30 states within seven of the ten EPA regions in the country. The EPA Permit Compliance System (PCS) data base to be described below lists 418 separate sources of pollutant discharge. BOD, Total Suspended Solids (TSS), and the pH levels of discharges are the three main conventional pollutants controlled, although in recent years Congress has initiated new regulatory efforts to also control toxic pollutants.

If EPA water pollution standards were set in the same manner as seat belt regulations or OSHA standards, a description of the regulatory constraints would be straightforward. In the seat belt and OSHA cases, firms face well-defined requirements on the technology or work environment. All firms must comply with the same set of regulations, such as ensuring that punch presses have the specified guards. There has been little change over time in the nature of the standards, except that some new regulations have been added. In contrast, EPA water pollution standards involve permissible pollution amounts that vary across firms and have varied over time.

---

[12] BOD is the standard measure of the organic pollutant content of water.

The 1972 Federal Water Pollution Control Act Amendments set the framework for regulation of industrial water pollution. The Act required that all sources discharging into the navigable waters of the country meet discharge standards based on the application of the "best practicable control technology" (BPT) by July 1, 1977, while complying with standards based on the "best available technology economically achievable " (BAT) by July 1, 1983.

In 1977 the Act was amended again, pushing back the 1983 deadline to July 1, 1984 and substituting a more complicated requirement. Conventional pollutants such as BOD and TSS were to meet standards based on the adoption of the best conventional technology (BCT), while toxic pollutants were to meet standards based on the best available technology (BAT).

The final BPT and BAT standards for various subcategories of the pulp and paper industry were promulgated on three separate dates: May 9, 1974, May 29, 1974, and January 6, 1977. The final BCT standards were issued on December 17, 1986 and left the BPT standards for BOD control unchanged. The BPT standards generally set limitations on the quantities of BOD that a plant could discharge per pound of pulp or paper produced.[13] However, the allowable discharges of BOD from each source were derived by multiplying this effluent limitation by the number of pounds of pulp or paper produced per day at the plant. This latter number formed the basis of the National Pollutant Discharge Elimination System (NPDES) permit required of each discharger. Since our empirical study covers the period from the first quarter of 1982 through the first quarter of 1985, the NPDES permits restricting BOD discharge were based on the 1977 BPT standards.

---

[13] For a formal description and analysis of the BPT rulemaking process, see Wesley A. Magat et al., Rules in the Making: A Statistical Analysis of Regulatory Agency Behavior (1986).

EPA possesses the authority to issue the NPDES permits, but in the case of 37 states that have met specified federal criteria the authority has been delegated to the states. States certified to issue NPDES permits also assume responsibility for their enforcement, which means inspecting the plants and taking action against sources found to be out of compliance. For states not certified to run their own permit systems, EPA issues and enforces the permits.

An important aspect of the permit process should be emphasized. EPA and the states do not set uniform permit levels irrespective of the industry characteristics associated with the pollution source. Each standard is industry-specific and represents pollution levels that are potentially achievable with available technologies. While these permits need not pass a benefit-cost test, an effort is made to ensure that a partial affordability criterion has been met.

Each source must regularly measure its pollution discharge levels and report its actual discharges of each pollutant in its permit on a monthly basis through a Discharge Monitoring Report

(DMR). If a source is out of compliance with the effluent standards in its permit, it is also required to file a non-compliance report. The states and EPA regional offices send the DMRs to EPA, which enters them into the PCS data base to serve as a basis for tracking compliance. In addition, EPA requires that Quarterly Non-Compliance Reports (QNR) be filed by each state and region each quarter to identify sources out of compliance. In the empirical study that follows, we will use the reported BOD discharge levels in the DMRs to measure the effects of inspections on BOD discharge levels.

Because the sources are required to report their pollutant discharge levels on a monthly basis, the on-site inspections play a somewhat different role than inspections carried out by other regulatory agencies, such as an OSHA inspection of an industrial site. The latter inspections constitute the primary basis for the agency to check compliance with its regulations and to have a visible presence in the workplace. In contrast, EPA or state-run inspections of industrial water pollution sources create a similar visible presence, but they provide only a secondary source of information about compliance because the monthly DMRs address the compliance question directly. Some NPDES permit inspections do test whether the DMR discharge levels are reported accurately and honestly, and they provide an incentive for firms to submit DMRs more frequently.

The difference between EPA inspections and OSHA inspections has also been narrowing years. Although the Bureau of Labor Statistics does not release the mandated injury reports to OSHA for compliance purposes, OSHA now gathers this information through on-site records checks to target its inspections. This procedure represents a partial and more time-consuming variant of the DMR process. Firms with good injury records are exempt from OSHA inspections.

The EPA inspections directly address one or more of the following items: the existence of an up-to-date permit, the installation of the abatement equipment necessary for compliance with the permit, management plans and practices, the preparation and maintenance of records, the correct operation of the abatement equipment, and the conduct of sampling and sample analysis. As a recent EPA report to O.E.C.D. explains, "Despite widespread self-monitoring, inspections remain the backbone of agency compliance monitoring programs....inspections are the government's main tool for officially assessing compliance, and for assuring quality control and lending credibility to self-monitoring programs. The independent evaluation provided by a government inspection is the key".[14]

EPA carries out three main types of inspections -- compliance sampling inspections, compliance evaluation inspections, and performance audit inspections. Compliance sampling inspections require approximately 30 work days of time to complete and involve actual sampling of the effluent at the plant, as well as an examination of the company's record-keeping system, its testing procedures, and its treatment system. In contrast, the compliance evaluation inspections take only about 3 workdays to complete. They involve no sampling, but the inspectors do examine the company's treatment facilities, monitoring methods, and records. The performance audit inspections require about 12 days of time to complete and consist of the same practices used in the performance evaluation inspection, plus observation of the permittee going through all of the steps in the self-monitoring process from sample collection and flow measurement through laboratory analyses, data work-up, and reporting. In addition, the performance audit inspector may leave a check sample for the permittee to analyze.

---

[14] Wasserman, supra, note 9, page III-7.

Based on the discharge reports in the DMRs and in the QNRs, as well as on the findings of inspections, EPA or the designated state agencies take enforcement actions against violators. Informal actions include telephone calls, warning letters, and notices of violation, as well as inspections. If these measures do not achieve the intended results, the control agencies can proceed with formal actions such as administrative orders, permit revision, formal listing of companies as ineligible for government contracts, grants, and loans, and finally, civil and criminal judicial responses.

Court action is a lengthy process involving the Justice Department that is only started as a last resort. Under Section 309(e) of the 1977 Clean Water Act, civil penalties could be awarded up to a level of $10,000 per day, while criminal penalties could range from $2,500 to $25,000 for the first violation and up to $50,000 for the second violation.[15] In addition, first violations could lead to imprisonment up to one year, with up to two years of imprisonment for the second violation. During the period from January 1, 1975 to July 1, 1985, EPA commenced only 64 judicial actions in the pulp and paper industry. Of these 42 cases resulted in fines and 4 were still pending at the end of the period. The fines varied from $1,500 to $750,000, with an average of $89,437. Because the regions lacked the incentives to regularly report enforcement actions other than inspections into the PCS data base, we concentrate our study on the effectiveness of the inspections on bringing firms into compliance with their permits. Thus, the inspections variable is intended to be a proxy for the overall enforcement effort associated with an inspection and all subsequent enforcement actions.

III. The Sample and the Variables

The Data Base

The PCS data base, which we utilize in our analysis, contains a file of inspections carried out on sources in every industry. The file for the pulp and paper industry shows only performance sampling and performance evaluation inspections for sources listing BOD discharge measurements, although the agency did carry out several performance audit inspections for lead pollution. For the period from the first quarter of 1982 through the first quarter of 1985 there were 276 inspections in the industry, of which 43 percent were compliance sampling and 57 percent were compliance evaluation.

---

[15] Under the 1987 Amendments to the Clean Water Act, the maximum civil penalty rose to $25,000 per day and the maximum criminal penalty increased to $50,000 for the first violation and $100,000 for the second violation.

The PCS data base lists 418 separate sources in the pulp and paper industry in its Inspections File, but only 77 of those sources submitted DMR measurements for BOD discharge into the Measurements File. The rest of the sources either did not submit DMRs including BOD measurements during the period under study, or their DMRs were not entered into the PCS data base, or they discharged pollutants other than BOD. Thus, we restrict our analysis to those 77 sources for which we have data on both inspections and BOD discharges. These 77 sources are located in six of the ten EPA regions. They are contained in SIC 26, further divided into five 4-digit SIC codes (2611, 2621, 2631, 2648, and 2661).

In this analysis we use calendar quarters as the unit of analysis. Only rarely was there more than one inspection for a given source in the same quarter. Despite the requirement that sources report DMRs every month to the state enforcement agency or EPA, which were then required to enter these data into the PCS data system, some DMR measurements are missing for the sources in our sample. In constructing the quarterly BOD measurements for our statistical analysis, we interpolated to fill in missing values and used averages of the BOD discharge levels within a quarter as the quarterly average BOD discharge levels.

Although the EPA analysts to whom we talked were confident that most of the discharge measurements in the DMRs were reported accurately, permittees do have several opportunities to cheat. They may choose not to report discharge measurements during months with unusually high discharge levels. This behavior would lead to some smoothing of the pattern of reported discharges, eliminating the top end of the distribution. More active attempts to mislead EPA include altering the contents of the sample being tested, falsely calibrating the test instruments, and recording false measurements in the DMRs.

Despite these possibilities for sending EPA misleading or false DMR discharge statistics, there are several incentives to report honest information in the DMRs. EPA follows the policy of attempting to inspect all major sources at least once a year. Compliance sampling inspections would detect whether most of the reported measurements were inconsistent with the measurements from the inspections, but they could not detect whether outliers were removed from the reports. Compliance evaluation inspections would detect the absence of the required abatement equipment, but would be less useful in evaluating whether the abatement systems were being operated correctly. Of course, the penalties for non-compliance and fraud in reporting also create incentives for truthful reporting of discharge measurements. The possibility of leaks to EPA by disgruntled employees makes this last incentive more compelling to firms considering manipulating their DMR data.

Taking into account the possibility that the DMR measurements may measure true compliance status with some error, it is still intructive to ascertain how well firms comply with the effluent regulations. Recently, the Environmental Protection Agency[16] issued a study of compliance by all the major pulp and paper mills (SIC 2611, 2621, 2631) in the eight Southeastern states comprising EPA Region IV over the period from the second quarter of fiscal year 1982 through the first quarter of fiscal year 1984. Eighty-two percent of the measurements fell within the permitted bounds. This compares with 75 percent of the measurements from the pulp and paper firms in our sample being in compliance. EPA further defines significant non-compliance for BOD as violations of the monthly average permit limits for any two months in a six-month period that exceed the limit by 40 percent, or violations of the monthly average limits for any four months in a six month period. Using this definition, 94 percent of the measurements indicated discharge levels not in significant non-compliance. The study also showed that four out of the 56 mills created most of the instances of significant non-compliance.

Sample Characteristics

Table 1 summarizes the means and standard deviations for the sample of the variables used in our analysis. The sample is a pooled time series and cross section of 77 plants followed on a quarterly basis from 1982:1 to 1985:1. The first two variables represent the pollution outcome measures that will be of primary interest as dependent variables in different equations. The

---

[16] U.S. Environmental Protection Agency, Study of Pulp and Paper Industry in Region IV (1986).

TABLE 1: Means and Standard Deviations of Variables Describing 77 Plants in Sample (1982.1 - 1985.1)

Variable	Mean	Standard Deviation
MQAVG (pounds/day)	5758.288	8919.173
MVIO (1 = out of compliance)	0.252	0.434
IQTR1 (1 = inspection 1 quarter prior to measurement)	0.248	0.432
IQTR2	0.273	0.446
IQTR3	0.273	0.446
IQTR4	0.281	0.450
IQTR5	0.300	0.458
IQTR6	0.295	0.456
REGN1 (1 = source located in Region 1)	0.095	0.293
REGN2	0.002	0.039
REGN3	0.064	0.244
REGN4	0.154	0.361
REGN5	0.039	0.193
REGN6	0.435	0.496
REGN7	0.000	0.000
REGN8	0.000	0.000
REGN9	0.000	0.000
REGN10	0.213	0.410
SIC11 (1 = pulp mill)	0.241	0.428

TABLE 1
(cont.)

Variable	Mean	Standard Deviation
SIC21 (1 = paper mill excl. building)	0.432	0.496
SIC31 (1 = paperboard mill)	0.253	0.435
SIC47 (1 = sanitary paper products)	0.012	0.111
SIC48 (1 = stationary products)	0.014	0.117
SIC49 (1 = converted paper)	0.000	0.000
SIC61 (1 = building paper or paper board mill)	0.048	0.214
TONS (daily output rate)	794.156	587.083

variable MQAVG is a continuous measure of the extent of pollution. It measures the number of pounds of BOD discharged per day, where this amount is averaged over the quarter. Although the amount of pollution is a variable of substantial economic interest, it is not the sole variable of concern. Different firms may have different permitted pollution levels so that, for example, a large plant may be in compliance with a high BOD level whereas a small plant may be in violation of its permit even though its discharge is less. Analyzing the effect of inspections on total discharges is, however, one of the most important ways of assessing the benefits of EPA's regulatory enforcement.

The second pollution variable, MVIO, is a discrete 0-1 variable that takes on a value of 1 if the pollution source is in non-compliance with its BOD standard in any of its monthly measurements in that quarter. This variable best captures whether the firm's performance is in compliance with its water pollution permit, but it does not reflect the extent of non-compliance. Unfortunately, it is not possible to construct a reliable measure pertaining to the amount pollution in excess of the permitted amount since data pertaining to the level specified in the permit are not available from the PCS data base. Instead, we will be restricted to MQAVG and MVIO rather than a hybrid of a continuous pollution measure and discrete compliance measure.

The next set of variables is a series of 0-1 dummy variables pertaining to whether the firm was inspected in a particular quarter. The variable IQTRJ is of the general form in which it takes on a value of 1 if the pollution source received an inspection J quarters previous to the pollution measurement in the current quarter, where J takes on a value from 1 to 6. It is quite striking that the rate of inspection is quite high, on the order of 25 to 30 percent of the quarters.

This relatively high inspection rate distinguishes the EPA enforcement effort from that of OSHA. Not only does EPA receive regular discharge monitoring reports from firms, but it also undertakes water pollution inspections at a rate of about one inspection annually per major pollution source. OSHA not only has no automatic data feedback mechanism, but it also has a much more sporadic inspections effort. In OSHA's early years, some analysts equated OSHA's inspection frequency to other rare events such as the annual chance of seeing Halley's Comet. At present, the OSHA inspection rate is much lower than this amount -- on the order of 1/200 for each firm in any year.[17] The intensity of EPA inspections consequently dwarfs that of OSHA inspections so that there is no reason to believe that the lack of efficacy of OSHA's minimal enforcement operation has any adverse implications for EPA's chances of success.

The variables of the form REGNJ are 0-1 dummy variables for the EPA region J in which the plant is located. These variables will be utilized to ascertain whether there are any important regional differences in pollution patterns. It should be noted that there are no pulp and paper mills located in three of the EPA regions (7, 8, and 9).

---

[17] See page 259 of W. Kip Viscusi, Reforming OSHA Regulation of Workplace Risks, in Regulatory Reform: What Actually Happened (L. Weiss and M. Klass eds. 1986).

The next set of six variables are of the form SICJK, which represents a dummy variable for the plant's four digit SIC industry code 26JK, where JK takes on the values 11, 21, 31, 47, 48, 49, and 61. Although all firms are in the pulp and paper industry, it was desirable to also include refined industry group dummy variables that reflect the firm's specific operations and technology. For example, pulp mills (SIC 2641) have different operations than converted paper plants (SIC 2649).

TONS is the final variable listed. It measures the number of tons of pulp and paper produced daily at the plant. Unlike the other variables in the data set, this variable was not included in the PCS data base. We matched each firm to a capacity measure using data provided in a published industry directory.[18]

IV. The Effect of Inspections on Pollution

The major purpose of this paper is to empirically measure the effects of inspections, along with their associated enforcement actions, on the behavior of firms in the pulp and paper industry. We will concentrate on an econometric approach which relates the conduct of an inspection in a given quarter to two measures of the firm's BOD abatement effort: (i) its absolute rate of effluent discharge (MQAVG); and (ii) whether its discharge rate falls below its permitted level (MVIO). As well, we also examine the effect of plant inspections on reducing the incidence of non-reporting of DMR data. To the extent that firms purposely refrain from reporting discharge levels during periods of non-compliance, the first two measures of the impact of inspections would be biased towards less impact than actually occurred. This third measure allows us to determine whether inspections improve the completeness of EPA's discharge monitoring system, which presumably leads to more discovery of non-compliance and, through subsequent enforcement efforts, further reductions in pollutant discharge levels.

---

[18] See Lockwood's Directory of the Paper & Allied Trades, (1983 ed.).

Empirical Framework for Measuring Abatement Effects

The underlying economic framework is straightforward, as pollution levels are governed by a capital investment process relating to the pollution control technology, as well as by the efficiency levels at which the abatement equipment is operated. The role of EPA inspections is to raise the expected cost of non-compliance, boosting the incentives for pollution reduction and compliance with the permit. Since the underlying theoretical basis is straightforward, we will proceed directly to the estimating equations.[19]

The equations to be estimated will be of the same general form whether the pollution variable is MQAVG or MVIO. To illustrate this general form, let $POLLUTION_{it}$ be the value of the pollution variable MQAVG or MVIO for pollution source i in period t. Some additional notation is needed before we can write down the equation to be estimated. The variable $IQTRJ_{it}$ is the 0-1 inspection variable for whether pollution source i was inspected in period t-J, $TONS_i$ is source i's capacity measure, SIC_i is a vector of four-digit SIC code dummy variables for pollution source i, $REGN_i$ is a vector of dummy variable for the EPA regions for source i, and $QUARTER_t$ is a vector of dummy variable for the quarters. The resulting estimating equation is of the form

---

[19] The model implicit here is articulated more fully for the analogous job safety case in W. Kip Viscusi, The Impact of Occupational Safety and Health Regulation, 10 Bell J. Econ. 117 (1979). More generally, see Richard Posner, Economic Analysis of Law, Third Edition (1986).

$$\begin{aligned} POLLUTION_{it} = \alpha + \beta_1\, POLLUTION_{it-4} &+ \sum_{k=1}^{n} \gamma_k\, IQTR_{t-k} \\ &+ \beta_2 TONS_i + \beta_3 SIC_i + \beta_4 REGN_i \\ &+ \beta_5 QUARTER_t + v_{it}, \end{aligned}$$

where v_{it} is a random error term. In the case of the continuous pollution measure, MQAVG, ordinary least squares is the appropriate estimator, whereas for the discrete compliance variable, MVIO, a logistic estimation procedure is employed. With some modifications, this equation is in the same general spirit as similar equations estimated for safety regulations.[20]

The first variable included is the lagged dependent variable, with the noteworthy distinction that the lag is 4 quarters rather than 1. The variable $POLLUTION_{it-4}$ is a proxy for the firm's stock of capital related to pollution control and for the general character of its abatement technology. Firms with high levels of pollution in the past are likely to continue to have high levels in the future because the nature of their control technology makes it costly to achieve pollution reductions. A four-quarter lag is utilized rather than a single quarter lag to capture the seasonality that often plays an important role in a firm's operations. The products produced, stream flow conditions, and the pollution permit amount may vary by season.

---

[20] The equation bears closest similarity to those in Viscusi, supra, note 8.

The lagged dependent variable serves an additional role with respect to regression-to-the-mean effects. It is possible that firms with an abnormally high pollution level in period t due to stochastic factors will be inspected in period t+1 and improve their performance compared with period t, wholly apart from any true inspection effect. Because the lagged values captures pollution levels, or compliance status, 4 quarters earlier, however, they are less susceptible to leading to inspection variable results that simply capture regression-to-the mean effects.

The next set of variables is a distributed lag on past EPA inspections. Evidence for OSHA suggests that there is generally a lag before firms can make the required capital investments to alter their performance level.[21]

Even if compliance only entails changes in operating procedures following an inspection, an effect may not be apparent until the next quarter. Consider a situation in which the firm files its DMR data for the first month in the quarter in the middle of the second month of the quarter. Even if EPA undertakes an inspection immediately, which is not usually the case, the sampling will not be completed until the middle of the final month of the quarter. Thus, under this best case scenario only half a month, or one-sixth of the pollution discharges for the quarter will be affected by the inspection. Because of the time lags before EPA receives the DMR data, the time needed before EPA can schedule an inspector to make a plant visit, the rather lengthy inspection process, and the time needed before EPA makes its report to the firm and the firm can take action upon it, no contemporaneous effect is expected.

---

[21] See Viscusi, supra, note 8.

Before requiring that any inspection effect enter with a lag, we tested empirically for whether the inspection variable led to a contemporaneous negative effect on pollution. Rather than observing a negative effect, there was a strong and statistically significant positive influence, which is consistent with the reverse causality hypothesis. We explored the causality issue in greater detail. Based on a Hausman[22] specification test, we were able to reject the hypothesis that the $IQTR_{it}$ variable is exogenous. Attempts to replace $IQTR_{it}$ (t=0) by an instrumental variable estimator also led to positive coefficients, suggesting that the primary relationship between the two variables is through high current levels of pollution leading to EPA inspections rather than inspections causing immediate reductions in pollution discharge levels. These results allow us to use only lagged inspection variables without losing any of the effects of the inspections on compliance, or creating a bias in our estimated coefficients.

---

[22] Jerry Hausman, Specification Tests in Econometrics, 46 Econometrica 1251 (1978).

The next variable, $TONS_i$, pertains to the capacity of the firm. Other things being equal, firms with larger capacity should produce more pollution MQAVG, but need not necessarily be more likely to be in or out of compliance with EPA standards. There may be economies of scale with respect to pollution control which would tend to make large firms less likely to be out of compliance. Similarly, the TONS variable may pick up factors related to the vintage of the technology to the extent that larger plants are newer and have less polluting technologies. If these large plants are considerably more efficient in controlling pollution, the absolute levels of pollution may be lower than smaller and more outmoded facilities.

Technological factors of this type will also be captured in the SIC code dummy variables, implying that differences in technologies and standards across parts of the pulp and paper industry will be taken into account. The regional dummy variables REGNJ also capture firm characteristics to some extent since plants in some regions tend to be older than those in other regions. These regional variables also reflect regional differences in standard setting and the nature of enforcement. These differences may be considerable due to the prominent role that the states have in the enforcement process.

The final set of variables is a series of 17 quarterly dummy variables for all but one of the quarters represented. This formulation was chosen over a simple time trend variable because of its greater flexibility. Not only do the QUARTER_t variables capture any possible uniform time trend, but they also capture other quarter-specific effects such as any seasonal and cyclical fluctuations production levels and water flows. Although some quarterly dummy variables were statistically significant, these coefficients are not reported since there was no apparent pattern evident in the results. We regressed MQAVG against both a continuous TIME variable and its square, but found no significant relationships.

Regression Results

Table 2 reports the ordinary least squares (OLS) results for the continuous pollution measure, MQAVG, and Table 3 reports the maximum likelihood estimates for the non-compliance variable, MVIO. Because of the close similarity of the findings, we will discuss each of the variables in turn for both of the tables.

The four-quarter lagged pollution variable has the expected strong positive effect on the current pollution status, which suggests that past pollution levels predict current discharge levels accurately because of the slowness of the capital expenditures process needed to transform their status. Since the MVIO variable has been altered by the logistic transformation, the results for the continuous pollution measure, MQAVG, can be interpreted more readily. It is quite striking that the weight placed on the four-quarter lagged pollution value is in excess of 0.98 in each of the four equations. Thus, there is almost complete replication of the pollution experience across time.

TABLE 2: Regression Equations for MQAVG (Quarterly Average BOD Discharge Levels in pounds/day)[a]

Independent Variables	Coefficents (Std. Errors) 1[b]	2	3	4
INTERCEPT	-434.029	-460.454	-494.034	-213.905
	(1683.935)	(1650.046)	(1592.309)	(1557.062)
MVIOT4	0.983	0.983	0.983	0.982
	(0.021)	(0.021)	(0.020)	(0.020)
IQTR1	-1174.689	-1059.423	-1064.031	-1148.911
	(517.225)	(511.525)	(497.787)	(487.430)
IQTR2	575.256	381.999	398.665	---
	(495.099)	(481.687)	(469.908)	---
IQTR3	-198.047	-155.912	---	---
	(467.133)	(463.305)	---	---
IQTR4	77.479	59.709	---	---
	(468.403)	(450.159)	---	---
IQTR5	374.924	---	---	---
	(468.248)	---	---	---
IQTR6	-584.136	---	---	---
	(440.411)	---	---	---
IQTR1x MVIOT1	---	---	---	---
	---	---	---	---
IQTR2x MVIOT2	---	---	---	---
	---	---	---	---
IQTR3x MVIOT3	---	---	---	---
	---	---	---	---
IQTR4x MVIO4	---	---	---	---
	---	---	---	---
IQTR5x MVIO5	---	---	---	---
	---	---	---	---
IQTR6x MVI06	---	---	---	---
	---	---	---	---
TONS	0.322	0.320	0.320	0.329
	(0.438)	(0.439)	(0.437)	(0.437)
SIC11	414.177	382.955	410.222	310.442
	(1408.440)	(1408.522)	(1394.943)	(1389.408)
SIC21	262.356	219.433	252.081	112.177
	(1418.355)	(1414.941)	(1393.285)	(1382.926)
SIC31	-205.950	-278.427	-253.484	-365.172
	(1426.645)	(1424.948)	(1410.265)	(1403.533)
SIC47	---	---	---	---
	---	---	---	---
SIC48	31.976	-41.814	17.162	-241.752
	(2806.433)	(2789.052)	(2719.955)	(2701.675)
SIC49	---	---	---	---
	---	---	---	---
REGN1	248.482	225.870	213.567	322.009
	(909.025)	(895.892)	(862.661)	(852.791)

TABLE 2
(cont.)

Independent Variables	Coefficents (Std. Errors)			
	1[b]	2	3	4
REGN2	---	---	---	---
	---	---	---	---
REGN3	-499.882	-500.628	-533.588	-360.471
	(1864.535)	(1823.241)	(1764.428)	(1751.873)
REGN4	230.897	219.572	204.147	310.894
	(890.368)	(846.406)	(807.680)	(797.493)
REGN5	59.067	107.966	115.888	147.361
	(1298.116)	(1299.413)	(1295.674	(1294.613)
REGN6	276.987	269.784	265.636	307.939
	(625.214)	(611.374)	(597.714)	(595.387)
REGN7	---	---	---	---
	---	---	---	---
REGN8	---	---	---	---
	---	---	---	---
REGN9	---	---	---	---
	---	---	---	---
Adj. R^2	0.903	0.903	0.904	0.904
N	373	373	373	373

[a] Each equation also included 17 quarterly dummy variables.

[b] Equation 1 uses a second-order polynomial distributed lag formulation fo IQTR1-IQTR6.

TABLE 3: Maximum Likelihood Equations for MVIO (Non-Compliance with BOD Standards)[a]

Independent Variables	Coefficients (Asymptotic Std. Errors) 1	2	3	4
INTERCEPT	-7.872	-7.648	-7.991	-8.012
	(23.008)	(22.884)	(22.884)	(23.113)
MVIOT4	2.650	2.637	2.640	2.641
	(0.362)	(0.359)	(0.356)	(0.356)
IQTR1	-1.12	-1.019	-0.920	-0.914
	(0.442)	(0.429)	(0.418)	(0.413)
IQTR2	-0.063	-0.134	-0.037	---
	(0.421)	(0.411)	(0.396)	---
IQTR3	-0.606	-0.644	---	---
	(0.398)	(0.396)	---	---
IQTR4	-0.030	-0.141	---	---
	(0.387)	(0.369)	---	---
IQTR5	0.448	---	---	---
	(0.389)	---	---	---
IQTR6	0.071	---	---	---
	(0.360)	---	---	---
IQTR1x	---	---	---	---
MVIOT1	---	---	---	---
IQTR2x	---	---	---	---
MVIOT2	---	---	---	---
IQTR3x	---	---	---	---
MVIOT3	---	---	---	---
IQTR4x	---	---	---	---
MVIO4	---	---	---	---
IQTR5x	---	---	---	---
MVIO5	---	---	---	---
IQTR6x	---	---	---	---
MVIO6	---	---	---	---
TONS	$-5.07x10^{-4}$	$4.971x10^{-4}$	$-5.127x10^{-4}$	$-5.124x10^{-4}$
	$(4x10^{-4})$	$(3.956x10^{-4})$	$(3.913x10^{-4})$	$(3.91x10^{-4})$
SIC11	6.321	6.263	6.396	6.405
	(22.998)	(22.875)	(23.108)	(23.106)
SIC21	5.800	5.754	5.958	5.968
	(22.999)	(22.876)	(23.109)	(23.107)
SIC31	5.352	5.306	5.423	5.431
	(23.00)	(22.877)	(23.110)	(23.109)
SIC47	---	---	---	---
	---	---	---	---
SIC48	2.506	2.404	3.064	3.084
	(23.077)	(22.951)	(23.178)	(23.175)
SIC49	---	---	---	---
	---	---	---	---

TABLE 3
(cont.)

Independent Variables	Coefficients (Asymptotic Std. Errors)			
	1	2	3	4
REGN1	1.709	1.791	1.540	1.531
	(0.746)	(0.736)	(0.690)	(0.683)
REGN2	---	---	---	---
	---	---	---	---
EGN3	2.188	2.474	2.033	2.015
	(1.481)	(1.412)	(1.336)	(1.319)
REGN4	1.098	1.316	1.101	1.094
	(0.685)	(0.655)	(0.621)	(0.615)
REGN5	1.835	1.868	1.951	1.950
	(0.888)	(0.889)	(0.877)	(0.876)
REGN6	-0.531	-0.404	-0.530	-0.535
	(0.524)	(0.505)	(0.482)	(0.480)
REGN7	---	---	---	---
	---	---	---	---
REGN8	---	---	---	---
	---	---	---	---
REGN9	---	---	---	---
	---	---	---	---
-2 Log L	281.30	282.64	285.35	285.39
N	374	374	374	374

[a] Each equation also included 17 quarterly dummy variables.

All else being equal (in particular, controlling for inspections), past pollution performance is close to a perfect predictor of current pollution levels.

The next set of variables pertains to the set of lagged inspection variables. Consider the continuous discharge measurements in Table 2. In equation 1 there is a second-order polynomial distributed lag over inspection variables for the preceding six quarters, equation 2 is a free-form lag over four quarters, equation 3 is a free-form lag over two quarters, and equation 4 includes only a single lagged value. The pattern is strikingly similar in all four equations. There is a consistently significant and substantial influence of IQTR on reducing discharge levels that occurs with a one quarter lag. Lagged values of more than a quarter are not consequential. The discrete compliance status equations in Table 3 convey the same influence of inspections, that is, they cause significant reductions in the rate of non-compliance in the subsequent quarter.[23]

The magnitude of the inspection effect is substantial. Consider equation 4 in Table 2. Each inspection reduces the value of MQAVG by 1149 pounds per day, which represents about a 20 percent reduction in the mean value of BOD discharges. Since the coefficients of subsequent IQTR variables are never significantly positive, there is no evidence of a significant post-inspection rebound in pollution discharge levels. These results imply that a permanent improvement in discharge levels takes place as a consequence of the inspection and all associated enforcement actions. Further, the 1149 pounds/day reduction in BOD in period t is reflected in an approximately equal reduction four quarters hence because the coefficient of MQAVG4 is 0.982. Thus, inspections substantially reduce BOD discharges after about one quarter, and they have a permanent effect on reducing the firm's future pollution levels.[24]

---

[23] We are uncertain why most of the effect of the inspections appears to occur in the quarter following the inspection. More detailed case studies of responses of individual plants to specific inspections would help explain this pattern of responses. Our results suggest that inspections tend to induce reduced discharge levels and enhanced compliance through immediate attention to better plant operation and maintenance, rather than longer term capital investments.

The compliance status results from Table 3 also indicate a large effect of the inspections, and their associated enforcement actions, on non-compliance rates. The coefficients of IQTR1 in equations 1 through 4 average -1.0, implying that had the source not been inspected its odds of being in non-compliance would have been about double. Since most plants in the sample were inspected about once a year and the average rate of non-compliance is 25 percent, the coefficients from the table suggest that without an inspection this non-compliance rate would have been 48 percent.

Finally, the TONS measure has the expected sign in each case, as firms with larger capacity have higher total levels of pollution and lower chances of being out of compliance. Neither effect is statistically significant, however. Similarly, the SIC and regional dummy variables fail to yield any statistically significant effects.

---

[24] A recent paper by Jonathan S. Feinstein (Detection-Controlled Inference, M.I.T. Department of Economics, 1986) provides an econometric argument for why the coefficients of the inspection variables would be biased downwards if detection of non-compliance through the non-submittal of DMR data, our results about the impact of inspections provide a lower bound on their true magnitude.

Effects on the Incidence of DMR Non-Reporting

While our econometric results in the beginning of this section clearly point to the conclusion that plant inspections cause firms to both reduce their pollutant discharge levels and come more closely into compliance with their discharge permits, inspections do serve other purposes as well. One of these is to induce firms to report more regularly their discharge levels to EPA or the designated state enforcement agency. We now examine whether inspections tended to reduce the incidence of DMR non-reporting.

Table 4 suggests that there is such a reporting effect. The first line in the table measures the difference between number of months that DMR data was submitted in the four months prior to an inspection and number of months with DMR data in the four months immediately following the inspection, averaged across all inspections in one of two periods, May, 1977 - November, 1984 and May, 1982 - November, 1984. The second line reports the analogous differences for a six-month period before and after the inspections, while the third line uses of 12-month period for DMR data. All six mean differences are negative and more than two standard errors away from zero, indicating high levels of

TABLE 4: Mean Difference Between the Number of DMR Reports Before an Inspection and the Number of DMR Reports After an Inspection

No. of Months of Possible DMR Data Prior To and After Inspection	Mean Difference Averaged Across All Inspections in Period (Std. error of mean)
1) Four months	
a) May, 1977-Nov., 1984	-0.386 (0.060)
b) May, 1982-Nov., 1984	-0.425 (0.108)
2) Six months	
a) July, 1977-Sept., 1984	-0.714 (0.090)
b) July, 1982-Sept., 1984	-0.868 (0.173)
3) Twelve months	
a) Jan., 1978-Mar., 1984	-2.107 (0.196)
b) Jan., 1983-Mar., 1984	-1.693 (0.477)

statistical significance. Thus, the completeness DMR reporting is clearly higher after inspections.

We must add one note of caution in interpreting these statistics because the mean differences are not adjusted for the trend of increased reporting of DMR data. Still, this trend could not explain much of this difference. To be conservative, consider the first line of the table reporting four months of DMR data where the trend ought to be least important. For both the long and short periods the mean difference averages about -0.10 reports per months, which implies that inspections cause one additional month of DMR data to be reported out of every 10 months. If the underlying trend of increased reporting accounted for, say, half of this difference (that is -0.05), then less than 20 months would have to pass before no more non-reporting of DMR data would occur. Since the period from May, 1977 to November, 1984 (line a) contains 84 months, the underlying trend must be negligible relative to the rates of increased reporting of DMR data implied by the mean differences in Table 4.

Thus, inspections do tend to cause increased reporting of DMR data, which in turn allows EPA to monitor more accurately, and therefore enforce, its water pollution standards.

V. Exploratory Benefit-Cost Analyses

One might conclude that EPA inspections are successful because all three of our measures of firms' responses to inspections show significant effects. From a social welfare perspective, however, this question requires valuing the benefits of the effluent reductions induced by an inspection and comparing these benefits to the full costs of each inspection. In what follows, we provide a preliminary exploration of the components of such a benefit-cost analysis. Unfortunately, the existing estimates of the benefits per ton of BOD eliminated per year are only approximate, and we could find no estimates of the compliance costs due to an inspection. As a result, this exercise is highly imprecise. Nevertheless, it does provide some perspective on the welfare consequences of the EPA inspection program for industrial water pollution.

Vaughan and Russell (p. 161)[25] have estimated the national benefits from the improvements in freshwater quality due to the BPT standards at $683 million (in $1980). While this estimate includes both the out-of-pocket expenses and the opportunity costs of the time of fishermen, it does not include the aesthetic benefits of fishing on cleaner waters, nor other benefits such as those from swimming and boating. Development Planning and Resource Associates[26] estimated that the BPT standards would reduce BOD discharges by 3,390,233 tons per year, which together with the previous estimate implies an average value of benefits per ton of BOD removed due to the BPT standards of $201.46.

Using equation (4) in Table 2, each inspection will tend to cause a reduction in BOD discharges of 1,148 pounds per day, or

---

[25] William J. Vaughan & Clifford S. Russell, Freshwater Recreational Fishing: The National Benefits of Water Pollution Control (1982).

[26] Development Planning and Resource Associates, Inc., National Benefits of Achieving the 1977, 1983, and 1985 Water Quality Goals (1976).

209.51 tons per year. Given the previous benefits estimate of $201.46 per ton, this implies that an average inspection produces $42,208 of benefits every year.[27]

Given the 0.982 coefficient of the MQAVG variable lagged four quarters in equations (4) in Table 2, the effectiveness of an inspection in maintaining lower effluent discharge levels decays at a negligible rate. Accepting the linear form of the equation and rounding this coefficient to 1.0, the equation implies that any BOD reductions from an inspection remain in force for years after the inspection. Thus, we can approximate the annualized benefits per inspection at about $42,208.

Given the mix of inspections in our sample of 43% compliance sampling inspections (requiring 30 days) and compliance 57% evaluation inspections (requiring 3 days), an average inspection required 14.6 days. Assuming the full cost of inspectors to be $50,000 per year over 220 working days yields a cost of $227 per day, or $3,315 per inspection. Annualizing this inspection cost at a ten percent discount rate gives an annual cost of $332.[28]

---

[27] This calculation assumes that the average benefits of each pound of BOD removed due to an inspection equal the nationwide average benefits of the BPT standards. This simplifying assumption ignores the fact that the effluent reductions at some plants induced by inspections will yield benefits much greater than the average, whereas inspections at other plants, even if they result in lower emissions, will improve water quality much less than for an average inspection. Without more disaggregated information about benefits, we were forced to make this simplifying assumption.

[28] The use of a ten percent discount rate is required by the Office of Management and Budget, but other more realistic rates would not significantly affect our conclusions.

Netting this cost of the inspection from the benefits gives an adjusted annualized benefit of $41,876 per inspection.

Consider now whether the annualized compliance costs incurred due to inspections are likely to exceed $41,876 per inspection. Since 75 percent of the firms sampled were already in compliance, we would expect them to spend little or nothing after an inspection. Thus, each non-complying firm must spend at least four times $41,876, or $167,504 per year, in order that the costs associated with a inspection exceed their benefits.

Whether compliance cost exceed this threshold probably hinges on whether the firm must make a capital investment to attain compliance, or whether a change in operating procedures will suffice. Although detailed cost data are not available for all portions of the pulp and paper industry, some suggestive statistics are available with respect to the costs of an activated sluge treatment system used to comply with the BPT standards in the waterpaper-molded products subcategory of the industry.[29]

For concreteness focus on the intermediate plant size (45 kg/day). Compliance for these firms entails an annual operation and maintenance outlay of $113,000, annual energy cost of $19,000, and an annualized capital cost of $339,000, leading to a total annual cost of $471,000.[30] If compliance following an inspection involves only the operation and maintenance costs, the expenditure of $132,000 is somewhat below the value of benefits less inspection costs. However, if a capital investment is required, the costs exceed the pollution reduction benefits net of enforcement costs by a factor of almost three.

---

[29] While the wastepaper-molded products subcategory is only one of many subcategories in the industry, the activated sludge treatment system represents a standard technology for biological treatment of pulp and paper mill wastes.

For small plants, with a total annual compliance cost (including amortized capital costs) of $288,000, the compliance costs outweigh benefits once capital costs are included. For large wastepaper molded-products plants with annualized compliance costs of $879,000, even the operation and maintenance costs of $176,000 exceed the pollution reduction benefits.[31]

To the extent that this particular case reflects the costs and the benefits for other subcategories of the pulp and paper industry, the following conclusion holds. If inspections lead firms to make substantial capital investments, then the costs of compliance exceed the benefits. Once having made these investments, firms may be more likely to undertake the appropriate operating procedures to maintain its compliance status as a result of an inspection. This promotion of continued vigilance on the part of firms that have already made the required capital investment is more likely to pass a benefit-cost test.

---

[30] All the cost estimates are found in U.S. Environmental Protection Agency, Development Document for Effluent Guidelines and Standards for the Pulp, Paper, and Paperboard and the Builders' Paper and Board Mills (1982).

[31] See U.S. E.P.A., supra, note 9.

VI. Conclusion

Compared with other health, safety and environmental regulations, EPA water pollution regulations for the pulp and paper industry represent an unusual success story. EPA sets standards for which compliance is feasible and then enforces these standards relatively vigorously, with inspections averaging over one per year for our sample. This mix is the opposite of that of OSHA, which has stringent standards coupled with weak enforcement. The coupling of regulations for which compliance is feasible with stringent enforcement is likely to create strong incentives for compliance, and the available evidence bears this out. There is a strong effect of inspections, and their associated enforcement actions, on both pollution levels and rates of compliance with the permit levels. In addition, inspections are associated with less non-reporting on pollutant discharge levels. Judged with respect to its legislative mandate to improve water quality, this effort is clearly a success.

One might raise the more general issue not treated by EPA's enabling legislation of whether the benefits accruing from this pollution reduction are commensurate with their costs. This calculation is in substantial need of better data to refine it, but some preliminary observations are in order. If one includes only the operation and maintenance cost associated with pollution control, then the benefits of inspections may exceed their costs. If capital costs are included as well, the results are probably reversed. One major difficulty associated with this calculation is that we cannot distinguish which incremental pollution control expenditures are associated with the effect of the inspections. Notwithstanding these caveats, it appears that the EPA water pollution regulations represent a dramatic departure from the apparent impotence of most other forms of health, safety, and environmental regulation. Remaining challenge is to set standards at a level that will ensure that the regulations are in society's best interests.